EXAMINING LONGITUDINAL CHANGE IN STUDENT TALK IN SMALL-GROUP LITERATURE DISCUSSIONS

Brendan David Hendrick

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in School Psychology in the School of Education

Chapel Hill
2017

Approved by:
Jeffrey A. Greene
Rune Simeonsson
Sandra Evarrs
Stephen Knotek
P. Karen Murphy
ABSTRACT

Brendan Hendrick: Examining Longitudinal Change in Student Talk in small-group Literature Discussions
(Under the direction of Dr. Jeffrey A. Greene)

Talk in classrooms has long been considered an avenue for the support of high-level critical thinking and comprehension. Quality Talk is one approach to small-group talk about and around literature that has shown promise in this regard. My study investigated changes over time in student talk during the Quality Talk approach. Two fourth grade classrooms participated in approximately 38 weeks of Quality Talk. Students were separated into heterogeneous groups and participated in Quality Talk discussions bi-weekly. Discussions were videotaped and professionally transcribed for analysis. Each transcript was prepared and student talk variables were counted. The variables mean length of utterance, words spoken per minute, and number of turns taken per minute by each student were investigated using longitudinal multi-level analysis. Results showed that the mean length of utterance varied significantly over time, words per minute spoken did not change, and turns per minute had a linear, upward trajectory. Mean length of utterance, words per minute, and turns per minute differed by text type, with students speaking more about narrative texts. Student oral reading fluency (ORF) was positively related to initial status in mean length of utterance and words per minute. Gender was positively related to initial status in turns per minute. Group assignment was found to be associated with change over time in all three talk variables. Student talk relative to their group was investigated with descriptive statistics. It was hypothesized that students would speak more similar length of utterance over
time, however, no clear pattern was apparent. These results add to the evidence around how talk outcomes in Quality Talk small-group literature discussions are related to student characteristics, text characteristics, and group assignment.
For Grandpa
This dissertation could not have been completed without the support and dedication of a few key people I have had the pleasure of learning from. First, I want to express my deepest gratitude to Dr. Jeffrey Greene for his measured, thoughtful support and ceaseless dedication to the power of scientific inquiry. I am also equally indebted to Dr. P. Karen Murphy, Dr. Carla Firetto, and all of the incredible professionals and scholars at Pennsylvania State University for their dedication to the Quality Talk project. In addition, I am humbled by the support and guidance I have received from the school psychology faculty, Dr. Steven Knotek, Dr. Sandra Evarrs, and Dr. Rune Simeonsson. From the first day onward, you have always instilled in me the deep belief that a school psychologist can be both a rigorous scientist and an effective practitioner. I am also indebted to my colleagues on the Quality Talk team at the University of North Carolina, Chapel Hill. Without our long conversations in the lab, I never would have developed the concept that became this dissertation. Lastly, thank you to my mother and father for raising me to question everything, to never stop learning, and for teaching me the value of quality talk.
TABLE OF CONTENTS

LIST OF TABLES.......................................................................................................................... xii
LIST OF FIGURES .......................................................................................................................... xiii
LIST OF ABBREVIATIONS .............................................................................................................. xiv

CHAPTER 1: INTRODUCTION ........................................................................................................ 1
  Quality Talk .................................................................................................................................. 6
  Purpose of Study ............................................................................................................................ 8
  Research Questions ..................................................................................................................... 10

CHAPTER 2: REVIEW OF LITERATURE ......................................................................................... 11
  Introduction .................................................................................................................................. 11
  Theoretical Basis .......................................................................................................................... 13
    Constructivist theory .................................................................................................................. 13
    Social constructivist theory ....................................................................................................... 14
    Summary of theories .................................................................................................................. 18
  Cognitive Model of Comprehension ............................................................................................ 19
    Summary of comprehension model .......................................................................................... 22
  Research on Classroom Discussion ............................................................................................. 23
    Current definitions of discussion .............................................................................................. 24
    Describing classroom interactions ............................................................................................ 24
    Understanding effective classroom interactions ....................................................................... 26
Changing classroom interactions ................................................................. 27
Discussions Designed for Reading Comprehension .................................. 30
Historical context ...................................................................................... 31
Discussion stances .................................................................................... 32
Whole group text-based discussion .......................................................... 34
Reciprocal Teaching .................................................................................. 35
Questioning The Author .......................................................................... 36
Paideia Seminars ...................................................................................... 37
Philosophy for Children ........................................................................... 40
Summary of whole group approaches ...................................................... 41
Small group discussions for reading comprehension ............................. 42
Grand Conversations ................................................................................ 43
Book Club ................................................................................................. 44
Literature Circles ...................................................................................... 46
Instructional Conversations ...................................................................... 47
Collaborative Reasoning .......................................................................... 49
Summary of small-group approaches ...................................................... 52
Quality Talk ............................................................................................... 53
Factors That Influence Student Talk ....................................................... 55
Expectation states theory ......................................................................... 56
Reading ability ......................................................................................... 57
Gender ....................................................................................................... 60
Text type ................................................................................................. 63
Summary of factors that influence student talk ................................................................. 65

Summary of Relevant Literature ......................................................................................... 65

CHAPTER 3: METHOD ....................................................................................................... 68

Participants ......................................................................................................................... 68

Procedures ........................................................................................................................ 69

Student experience of Quality Talk .................................................................................. 70

Variation in group assignment ......................................................................................... 73

Video recordings ................................................................................................................ 75

Oral reading fluency .......................................................................................................... 76

Student and text characteristics ....................................................................................... 76

Data Preparation ............................................................................................................... 77

Non-word and inaudible speech ....................................................................................... 77

Beginning of relevant discussion ..................................................................................... 77

Defining turns .................................................................................................................... 78

Interruptions ...................................................................................................................... 79

Variables ........................................................................................................................... 80

Variables for research question 1..................................................................................... 82

Variables for research question 2..................................................................................... 83

Variables for research question 3..................................................................................... 83

Variables for research question 4..................................................................................... 84

Data Analysis .................................................................................................................... 84

Multi-level longitudinal analysis ..................................................................................... 85

Advantages of longitudinal multi-level models ............................................................... 86
LIST OF TABLES

Table 1 – Descriptive statistics of participants.................................................................69
Table 2 – Descriptive statistics of participants by group.................................................75
Table 3 – Quality Talk discussion weeks and text types..................................................81
Table 4 – Variables and analysis methods by research question...................................85
Table 5 – Descriptives of Level 1 and Level 2 variables ..............................................93
Table 6 – Descriptives of Level 1 variables, by group..................................................94
Table 7 – Correlation Matrices......................................................................................95
Table 8 – Teacher words/turns by group and time point..............................................96
Table 9 – p-values of group comparisons for MLU.......................................................105
Table 10 – p-values of group comparisons for Words/Min...........................................109
Table 11 – p-values of group comparisons for Turns/Min............................................111
Table 12 – Summary of RQ 3 variables.......................................................................125
LIST OF FIGURES

Figure 1 – Frequency of teacher questions by time point………………………………………..73

Figure 2 – Standard deviation of group utterances …………………………………………..112

Figure 3 – Standard deviation of group utterances, narrative stories………………..113

Figure 4 – Standard deviation of group utterances, informational stories……………114
LIST OF ABBREVIATIONS

AT: Accountable Talk
ANOVA: Analysis of Variance
BC: Book Clubs
CR: Collaborative Reasoning
CI: Construction-Integration Theory
ELL: English language learners
EST: Expectation States Theory
GC: Grand Conversations
IC: Instructional Conversations
IES: Institute of Educational Sciences
LMLM: Longitudinal Multi-Level Model
MLM: Multi-Level Model
NAEP: National Assessment of Educational Progress
NCLB: No Child Left Behind
OLS: Ordinary Least Squares
QT: Quality Talk
QtA: Questioning the Author
P4C: Philosophy For Children
PS: Paideia Seminars
RT: Reciprocal Teaching
TT: Thinking Together
CHAPTER 1: INTRODUCTION

American students continue to struggle to achieve high-level comprehension of literature. Results of the recent National Assessments of Educational Progress (NAEP; National Center for Educational Statistics, 2013) indicated that most students in the United States did not demonstrate high-level comprehension in reading. Specifically, in 2013 only 8% of fourth graders were assessed at the advanced level, 27% at the proficient level, and 65% at the basic or below basic level. Over the past four years (2009-2013) reading achievement on the NAEP has remained largely unchanged. Although a small percentage of fourth grade students were measured at the advanced level across the nation, on a state by state basis, no state averaged above proficient in 2013 (National Center for Educational Statistics, 2013). These results are concerning as the NAEP utilizes the best available standards for proficiency in reading comprehension (Kamil, Afflerbach, Pearson, & Moje, 2011). According to the NAEP Framework, the highest level of reading comprehension (i.e., advanced) requires students to “...make complex inferences and support their inferential understanding of the text. Students should be able to apply their understanding of a text to make and support a judgment” (National Assessment Governing Board, 2012, p. 41). By contrast, the lowest level (i.e., basic) involves identification and recall of textual details (National Assessment Governing Board, 2012). It is apparent, therefore, that relatively few students were able to demonstrate deep understanding of text or make critical judgments of text as assessed by the NAEP.
Deficits in comprehension have long been of concern to educational scholars (Levin, 2004; Murphy, Wilkinson, Soter, Hennessey, & Alexander, 2009; Smith & Szymanski, 2013). As early as the 1970s, institutions such as Center for the Study of Reading at the University of Illinois conducted comprehension research, calling attention to the lack of comprehension instruction in typical classrooms (Kamil et al., 2011). More recently, increased attention has been given to teaching critical thinking, reasoning, and higher-order thinking skills (Chang-Wells & Wells, 1993; Murphy et al., 2009; Smith & Szymanski, 2013). Numerous scholars have attempted to codify high-level comprehension skills in order to guide classroom practices (Smith & Szymanski, 2013). Wells (1990) used the term “literate thinking” (p. 13) to describe not just the absorption of information, but engagement with text to enhance meaning between and within student’s minds. It has been emphasized that this “literate thinking” is what makes the “evolution of higher mental processes” (Chang-Wells & Wells, 1993, p. 62) possible. Murphy et al. (2009) used the term “critical literacy” (p. 741) to describe understanding that reaches beyond simple facts to deeper concepts. Regardless of the specific term used, it is clear that high-level comprehension is increasingly considered essential for student success (Levin, 2004; Trilling & Fadel, 2009).

Classroom discussion approaches are one potential solution to the problem of high-level comprehension instruction and deserve continued study. Discussion approaches have been both widely promoted and empirically shown to help students gain knowledge (Alexander, 2008; Henning, 2008; Mercer & Littleton, 2007; Nystrand, Gamoran, & Heck, 1993; Raphael, 1998; Reznitskaya, 2012; Tharp & Gallimore, 1988). In addition, learning through discussion has been empirically linked to educational outcomes such as increased comprehension and increased argumentation skills (Gillies, 2014; Murphy et al., 2009; Reznitskaya et al., 2001; Reznitskaya,
Kuo, Glina, & Anderson, 2009; Soter et al., 2008). Historically, the goals of classroom discussion have evolved alongside the goals of school itself (Murphy, Wilkinson, & Soter, 2011). For decades, discussion was only used to assess student knowledge via recitation of facts (Henning, 2008). As pedagogical theories developed, educational thinkers embraced the utility of discussion as an intellectually beneficial activity. Beginning in the early 20th century, proponents of classroom discussion developed theories about the potential benefits of children talking to each other productively (Harnack, 1968; Henning, 2008; Thayer, 1928). It was during this time that educational thinkers also began to promote the idea that students were active participants in creating knowledge for themselves (Piaget, 1932; Vygotsky, 1978). According to Dewey (1910), discussion allowed students to experience each other’s thinking and then internalize meaning for themselves (Harnack, 1968). Vivian Thayer (1928) wrote that discussion between students supported engagement, intellectual risk taking, and development of thinking skills. These concepts have influenced the goals of both discussion research and education at large.

In the contemporary educational landscape, national curriculum standards include classroom discussion. The Common Core State Standards include multiple standards stipulating discussion in some form (National Governors Association Center for Best Practicies & Council of Chief State School Officers, 2010). For example, within the Kindergarten–12th grade English Language Arts standards, the Common Core has a section devoted exclusively to standards for Speaking and Listening. Beginning at first grade, Common Core standards include language stating that students should know how to follow participation rules, build on others’ questions, and ask questions of their peers in discussion (National Governors Association Center for Best Practicies & Council of Chief State School Officers, 2010). Other content specific elements of the Common Core also align with the use of discussion for high-level comprehension. For
example, the reading comprehension standards include integration of complex themes and the critical evaluation of arguments in a text. The integration of discussion into these standards is further evidence that effective discussion has a critical role in best practices for learning.

Furthermore, discussion is increasingly seen as an element of so-called “21st century” skills (Kuhn, 2015; Reznitskaya, 2012; Trilling & Fadel, 2009). These are skills and behaviors thought to best prepare children for the fast pace of economic development and technology in the coming decades (Trilling & Fadel, 2009). According to this perspective, children will be more likely to thrive in future professions if classroom teachers focused on skills like collaboration and communication (Murnane & Levy, 1996). Results of economic research on the U.S. labor market further support the importance of discussion. Deming (2015) noted that beginning in the 1980s there has existed a clear trajectory of employment growth in jobs requiring social abilities, both in the highest and lowest paying positions. Classroom discussion can promote these 21st skills, by supporting student development of critical-thinking and collaboration while building knowledge (Murphy et al., 2009; Reznitskaya et al., 2001; Soter et al., 2008). Overall, quality classroom discussion promotes the development of skills that will better prepare students for the jobs that are likely to be available when they enter the labor market (Deming, 2015; Trilling & Fadel, 2009).

Unfortunately, high quality discussion continues to be a sporadic occurrence in modern classrooms (Nystrand, Wu, Gamoran, Zeiser, & Long, 2003; Reznitskaya, 2012). Teacher controlled discussions have continued to be favored over discussion that supports critical analytic thinking and comprehension. Results of research by Nystrand et al. (2003) indicated a majority of classrooms studied did not feature open exchange of ideas, but rather teachers simply soliciting answers from students. Explanations for this dearth of engaging discussion in
classrooms have been offered by various researchers (Burbules, 1993; Henning, 2008). One proposed barrier was that teachers prefer a more “autocratic” (Henning, 2008, p. 3) teaching style. Others have noted that growing class sizes may make discussion less viable (Burbules, 1993; Henning, 2008). Also, though the proliferation of standardized assessments has accelerated research and turned attention towards higher-level comprehension, some have hypothesized that standardized testing has encouraged teachers to focus on rote facts instead of discussion (Henning, 2008). Teachers who want to integrate discussion practices often use a patchwork of various interventions without any systematic measurement of outcomes (Henning, 2008). Regardless of these proposed barriers, more research is needed to provide evidence that discussion can positively impact communication skills, high-level comprehension, and critical analytic thinking. This additional evidence will hopefully energize the implementation of quality discussion in more classrooms across the U.S.

In sum, national assessment data has indicated that a majority of students in the United States fail to meet benchmarks for high-level comprehension. Quality classroom discussion approaches have been shown by research to be a valuable potential solution to this issue. The use of discussion is supported by current national standards initiatives and provides invaluable practice in skills necessary for the economies of the future. In spite of proponents in research and professional communities, quality discussion has yet to become a common feature of classrooms (Nystrand et al., 2003; Reznitskaya, 2012). Discussion approaches have been developed and researched in an attempt to provide teachers with effective and feasible frameworks for achieving high quality classroom discussion. One such approach, Quality Talk, is an example of an evidence-based discussion approach that can be used to research how mechanisms of
discussion operate to enhance high-level comprehension, critical analytic thinking, and communication skills (Wilkinson, Soter, & Murphy, 2010).

Quality Talk

Quality Talk is a small-group literature discussion approach developed to bolster high-level comprehension and critical thinking through interaction about and around text (Wilkinson et al., 2010). This approach was developed in part as a result of a major meta-analysis of the empirical research on nine previously existing discussion approaches (Murphy et al., 2009). Quality Talk’s creators synthesized the most effective features of previous approaches to create an optimal framework for literature discussion (Wilkinson et al., 2010). The Quality Talk approach identifies student behaviors, teacher behaviors, and the underlying environmental conditions for high-quality discussion. These features were based on extensive reviews of effective discussion environments (Wilkinson et al., 2010). During the course of Quality Talk, students are taught discussion skills that bolster high-level comprehension. Each skill (i.e., high level questioning) was identified by empirical research to be strongly associated with comprehension (Wilkinson et al., 2010). Strategies used by Quality Talk teachers were also identified via research on which explicit “teacher moves” (Wilkinson et al., 2010, p.154) were most helpful in implementing high quality discussions. Concurrently, pedagogical principles that support discussion are also explicitly stipulated in the Quality Talk model. Adherence to these principles creates a culture of inquiry wherein discussion is the means by which students create knowledge. Overall, Quality Talk was designed to be a culmination of the most effective parameters and strategies for achieving student discussion that results in better comprehension of text (Wilkinson et al., 2010).
The theoretical concepts underlying Quality Talk further support its utility in producing high-level comprehension amongst students. From a social constructivist standpoint, Quality Talk features conditions for active engagement, scaffolding, and co-creation of meaning (Almasi & Garas-York, 2009; Vygotsky, 1978; Wilkinson et al., 2010). By questioning each other and providing explanations to peers, students in Quality Talk discussions can cooperatively experience deeper understanding of the text. From a constructivist standpoint, students in Quality Talk discussions are challenged to explain their perspectives and integrate contrasting ideas to create knowledge for themselves. By talking through ideas about stories, knowledge is formed by individuals as they accommodate new ideas (Piaget, 1932; van Blankenstein, Dolmans, van der Vleuten, & Schmidt, 2009). In addition, Quality Talk discussion is aligned with current models of how comprehension occurs in students. According to Construction-Integration (CI) theory of comprehension, comprehension occurs when students construct meaningful mental models of the text and then integrate this information into prior-held notions and beliefs (Kintsch & Kintsch, 2005). Quality Talk discussions make the process of comprehension transparent as students elaborate on their thinking.

Current research on Quality Talk is promising with regard to enhancing students ability to critically analyze text (Li, Murphy, & Firetto, 2014; Wilkinson et al., 2010). Detailed analysis of the quantity of individual student talk during Quality Talk has yet to be conducted, however. Specifically, it is unknown how individual students’ talk outcomes change over the course of participation in multiple Quality Talk discussions. It is also unknown if those changes are associated with gender or reading ability of participants. Student talk is thought to be critical to the theoretical mechanisms by which Quality Talk fosters high-level thinking; therefore more research is needed to understand how talk changes over the course of the intervention. From a
social constructivist standpoint, talk is considered a key mechanism by which students negotiate meaning and develop cognitive skills (Au & Mason, 1981; Baker, 2009; Chang-Wells & Wells, 1993; Palinscar, 1998). Thus, if a student’s talk does not increase in discussion, the student may not be engaging in the co-creation of meaning. From a constructivist standpoint, the presence of student talk in discussion suggests a student is engaging with new ideas to create new meaning for themselves (Mercer & Littleton, 2007; Piaget, 1932). A lack of increase in talk during the course of the Quality Talk intervention would indicate that discussions are not activating one important theoretical mechanism for bolstering high-level comprehension and critical analytic thinking (Palinscar, 1998).

**Purpose of Study**

The purpose of my study is to investigate the nature of change in student talk over a 38-week implementation of the Quality Talk small-group discussion approach. Quantitative methodologies can be utilized to shed light on how student talk changes and what student characteristics may predict these changes. In researching “talk,” I am referring to the utterances made by students in response to each other or the teacher in the context of a group discussion. My study of talk is situated, at least in part, within the tradition of sociolinguistic analysis often employed by educational researchers to reveal the transactions that take place in the learning environment (Cazden, 2001; Edwards & Westgate, 1987). This tradition of research on talk consists of various methods of analysis of spoken interactions, which is highly relevant to my study (Edwards & Westgate, 1987). I note this in order to contrast my study with research on “speech,” which carries with it the connotation of articulation skills, pragmatic skills, verbal fluency, or other features (Hegde & Maul, 2006). As defined by Hegde and Maul (2006), “speech” refers to “the actual production of oral language” (p. 6). Due to its focus on the
production mechanisms, generally the research on “speech” is less suited to the analysis of how students and teachers co-create meaning through language. Research on “speech” is of a more individual and medical nature (e.g., speech disorders, verbal motor disorders) than research on talk (e.g., classroom talk analysis, analysis of student-student interactions, measurement of turns at talk, etc.). Thus, I used the term “talk” in order to clarify both the scope and conceptual framework for my study.

Based on extensive theoretical and empirical investigations of talk, small-group literature discussions are assumed to result in more similar participation, enriched responses to text, and more participation. Through this study, I will investigate whether these assumptions regarding talk were born out in how much students talked across Quality Talk discussions. My study was not designed to assess if student talk increased in quality, however. Although this is an important question, my study will investigate changes over time in basic measures of talk (i.e., works spoken, turns taken) as a step towards a better understanding of the relationships between discussions and students. Although it is hypothesized that increases in talk will be found, my analysis will answer questions regarding the relationship of talk to genre, reading ability, gender, and group assignment. My study will add information to the research on Quality Talk and small-group discussions in general, as few studies have included data on talk on a student level. In fact, much of the existing research has averaged talk measures over a whole class or groups (e.g., Daniels, 2002; Eeds & Wells, 1989; Lipman, 1980). This has masked the observation of gaps in participation or differential effects that various sub-groups of students may have experienced during discussion interventions. The results of my investigation will help researchers and educators better understand how Quality Talk relates to changes in talk outcomes on the individual and group level.
Research Questions

To investigate student talk outcomes associated with Quality Talk small-group literature discussions, the following research questions were proposed:

RQ 1: How does student talk change over time during small-group literature discussions?

RQ 2: How is text type related to student talk during small-group literature discussions?

RQ 3: What student characteristics explain change in student talk over time?

RQ 4: How does student talk change relative to other group members over time?
CHAPTER 2: REVIEW OF LITERATURE

Introduction

Classroom discussions can take many forms, ranging from teacher-led whole groups to peer-led small-groups. In discussion students talk, share ideas, answer questions, and question each other (Wilkinson, 2009). Effective discussions have the potential to increase engagement with text, increase critical analytic thinking skills, and support high-level comprehension (Almasi & Garas-York, 2009; Keefer, Zeitz, & Resnick, 2000; Murphy et al., 2011; Wilkinson et al., 2010). Students are thought to benefit from discussions both cognitively and socially. Consequently, scholars have sought to better understand how these benefits occur and how to maximize them (Gillies, 2014; Murphy et al., 2011).

Research on classroom discussion is diverse in scope and methodology. It also has been conducted over decades during which educational and empirical priorities have changed. Critically, research priorities have shifted from describing phenomena to establishing experimental evidence for classroom interventions, including those designed to create effective discussion (Nystrand, 2006; Snow, 2002). Furthermore, recent educational standards prioritize high-level comprehension of texts and collaboration skills (Kuhn, 2015; Murphy et al., 2011; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). Advances in developmental theory have also shaped the expected outcomes and rationales for classroom research (Gillies, 2014; Webb & Palinscar, 1996). Educational funding priorities in the United States have accelerated interest in research to promote critical thinking.
To this end, the strategic use of talk is commonly promoted as way of promoting critical thinking in today’s classrooms (Keefer et al., 2000; Kuhn, 2015). As Kuhn (2015) pointed out, however, the evidence regarding students’ intellectual engagement with each other is inconsistent and discussion “does not always yield identifiable benefits” (p. 1). In light of these developments, more research on classroom discussion approaches for high-level comprehension is warranted.

My goal is to review the research relevant to small-group literature discussions highlighting changes in student talk associated with discussion. This goal includes establishing the research context for analysis of student talk outcomes and possible explanations of change in student talk. As research on discussions is informed by developmental theory, I will review relevant theories in order to contextualize the mechanisms by which discussions operate. Early research on classroom discussions demonstrated certain critical features and documented established patterns of interaction, thus informing subsequent research on teacher-student and student-student interaction (Cazden & Beck, 2003; Mehan, 1982; Nystrand, 2006). This research will be briefly reviewed, as well as research on approaches that promote higher quality classroom interaction, in general (Mercer, Wegerif, & Dawes, 1999; Michaels, O’Connor, & Resnick, 2008; Tharp & Gallimore, 1991). Although reading comprehension is one outcome among many that have been examined in the discussion research, I will specifically cover the research on student talk outcomes during reading comprehension discussions. Talk outcomes in whole group literature discussion approaches will be reviewed first, followed by research on talk outcomes in small-group approaches. Some student characteristics and the type of text may account for variation in student talk (Bonito & Hollingshead, 1997; E. Cohen, 1994; Li et al.,
2014; Webb & Palinscar, 1996). I will synthesize this research as well; highlighting gaps with regard to student talk outcomes. In the next few sections I will review theories of development that have guided research on literature discussions approaches. Constructivist theory will be discussed first. Due to the fact that discussions, by definition, involve interactions between people, social constructivist theory will be also be reviewed.

Theoretical Basis

**Constructivist theory.** Constructivist theory helps to describe the mechanisms by which discussion promotes conceptual change (Gillies, 2014; Murphy et al., 2011). Piaget (1932) theorized that learning occurs when students adjust their thinking to a new concept (Mercer & Littleton, 2007; Webb & Palinscar, 1996). This “accommodation” as it is called, is thought to be how students construct knowledge for themselves as individuals (Piaget, 1932). According to constructivist theory, the catalyst of learning is the conflict between a child’s prior schemas and the presentation of new schemas. When these new schemas come from others, it is a particular kind of conflict, termed “sociocognitive conflict” (Almasi, 1994; Webb & Palinscar, 1996, p. 844). With regard to this type of learning, Piaget (1976) wrote that interactions with adults were unlikely to lead to the same kind of conflict as peer interactions. His reasoning was that children were more apt to accommodate to their equals as opposed to a higher status adult. As an illustration of this, McKeown, Beck, and Blake (2009) found that students who used strategies with each other to construct meaning from texts learned more than students exposed to direct instruction from an adult. This is just one instance in which students were observed accommodating information from peers via interaction (McKeown et al., 2009).

From the vantage point of the individual, constructivist theory has guided research on which mechanisms explain or accelerate learning during classroom discourse (Murphy et al.,
2011). With these processes in mind (i.e., accommodation, sociocognitive conflict), interventions for highly effective classroom discourse can be designed and outcomes from these approaches can be measured (Murphy et al., 2011). Constructivist theory is not the only theoretical viewpoint from which research into small-group discussions can be situated. Social constructivist theory also has a significant place in the research on classroom discussions.

**Social constructivist theory.** Social constructivist theory has guided many principles of group learning, in general, including classroom discussions (Almasi & Garas-York, 2009, Palinscar, 1998). Social constructivist theory posits that the development of knowledge is a social process in which an individual creates new knowledge resulting from an interactive situation. The work of Russian thinker Lev Vygotsky (1978) is often cited as the basis for this theoretical framework. According to Vygotsky (1978), social interactions drive knowledge development and the cultural framework in which those interactions are situated is highly influential. In addition, Vygotsky thought of language as a cognitive tool, a social tool, and a pedagogic tool (Mercer et al., 1999; Vygotsky, 1978). These multiple functional perspectives on language have inspired research focused on interpersonal interactions as sites of learning and development (Palinscar, 1998). This stands in contrast to research that focused on interactions in classrooms as sociological events (e.g., Edelsky, 1981; Shultz, Florio, & Erickson, 1982).

To elaborate, from the social constructivist perspective, knowledge itself is thought to be the result of multiple actors engaged in mutually influential exchange (Baker, 2009). This concept was referred to by Baker (2009) as the “co-construction of knowledge” (p. 3). From this perspective, knowledge emerges from interaction where meaning is negotiated collaboratively with the participation of various contributors. Subsequently, the resultant collaborative knowledge can be difficult to separate from the knowledge of the individual as each individual
appropriates his or her own knowledge (Baker, 2009). Many discussion approaches are designed to both develop students’ higher-order thinking and advance students’ knowledge through teaching this negotiation of meaning, making social constructivist theory highly relevant to discussion research. Palinscar (1998) pointed out, however, that social relationships may sometimes work against group meaning making. For example, if a child feels they have lower status than the other group members, they may contribute less. Similarly, Nyikos and Hashimoto (1997) wrote that potential growth may be limited depending on whether group conditions are productive for learning. Even so, from a social constructivist perspective, coping with divergent perspectives has been noted as a key element in the development of critical thinking (Nyikos & Hashimoto, 1997). Vygotsky (1978) theorized that higher order critical thinking stemmed from social interaction and self-knowledge. Thus, comprehension would be achieved via an individual interacting with others and consciously thinking about his or her thinking (Almasi & Garas-York, 2009; Vygotsky, 1978; Webb & Palinscar, 1996). According to this perspective, students must participate in interactive experiences to gain comprehension skills, while using metacognitive strategies to reflect on what they are constructing (Baker, 2009; Nyikos & Hashimoto, 1997; Palinscar, 1998). Interactions though talk have been seen as critical to the process by which higher-order thinking and comprehension are achieved (Nyikos & Hashimoto, 1997; Palinscar, 1998).

Pearson and Gallagher (1983) reviewed educational research and built on social constructivist ideas, resulting in the concept of “gradual release of responsibility” (p. 338). The gradual release of responsibility model was developed from a synthesis of research into effective classroom intervention in reading. They described a model in which a teacher guides the students towards independence work by removing support as the students gain skills (Pearson &
Gallagher, 1983). Although it shared some features with the explicit instruction model proposed by Rosenshine (1979), it was different in that it allowed for greater flexibility in classroom application. Many classroom discussion approaches are implemented with the gradual release of responsibility as an element of the intervention (Murphy et al., 2009; Soter et al., 2008). Without this idea, it would be unlikely that the genuine co-construction of mutually negotiated meaning could truly take place among the groups (Baker, 2009).

The concept of scaffolding has also influenced classroom discussion research and intervention. Wood, Bruner, and Ross (1976) first used this term to describe their research on how adults help children solve problems. As defined by Wood et al. (1976) scaffolding involves recruiting the participant in the task, making the task manageable, maintaining direction, controlling frustration, fading, and modeling solutions. It was observed that children learned more when tutors helped with a puzzle at just the right moment, based on that child’s proficiency (Wood et al., 1976). Key to this process is the fading of teacher support. In interventions design to support high-level comprehension of text though discussion, the fading of support takes place in part through intentional teacher moves (Wilkinson et al., 2010). Scaffolding and gradual release of responsibility are both frameworks for thinking about how children become more independent. Scaffolding, as defined by Wood et al. (1976), referred to a specific task or problem that needs support from a more skilled other. Gradual release of responsibility, as defined by Pearson and Gallagher (1983) described a particular model of effective educational interventions extrapolated from research, generally. Nevertheless, both concepts are often implemented to increase student independence. Due to the fact that a key goal of many classroom discussions is the development of students’ independent thinking, finding the correct level of teacher influence is critical. As Baker (2009) notes, the collaborative co-creation of meaning should be “genuine”
(p. 3), and independent student thought can be supported through these concepts when support is “faded” over time across experiences.

When specifically applied to discussion, scaffolding has been shown to take many forms during classroom discussions. Providing a few broad examples, Almasi and Garas-York (2009) articulated two categories of scaffolding that are present in classroom discussions, specifically. First they described “microgenetic scaffolding” (p. 474) in which support for high-level comprehension through discussion is calibrated on a moment-by-moment basis during the act of discussion. Examples of this support include a teacher helping a student form a question, asking an open ended question, summarizing an idea, or helping a student understand another student’s perspective (Almasi & Garas-York, 2009; Nystrand et al., 2003; Wilkinson et al., 2010). This kind of scaffolding is prevalent in discussion approaches that are teacher led (e.g., Questioning the Author, Beck, McKeown, Hamilton, & Kucan, 1997; Paideia Seminars, National Paideia Center, 2015; Collaborative Reasoning, Waggoner, Chinn Yi, & Anderson, 1995; Philosophy for Children, Lipman, 1980) and support reasoning, thinking, and development of high-level comprehension in a discussion environment. This type of scaffolding for students could also take place between the students themselves as they respond to each other’s reasoning as the discussion plays out (Almasi & Garas-York, 2009).

Almasi and Garas-York (2009) also noted that discussion approaches often include scaffolding before and after students interact. This “ontogenetic scaffolding” (p. 476) supports both a student’s preparation for interaction and reflection following interaction. It is in this way that teachers can influence a student’s perspective on the construction of knowledge (Almasi & Garas-York, 2009). For example, by brainstorming questions before a discussion, students can become more prepared to integrate new ideas. Students can also solidify the new knowledge they
were exposed to by peers through teacher-led reflection after discussion. In sum, effective ontogenetic scaffolding results in students who are primed to co-create knowledge from the ideas of others. This type of scaffolding is most prevalent in classroom discussion approaches that de-emphasize the teacher’s role in direct discussion leading. These approaches hope to create a setting in which students can fully experience the advantages of interacting with each other (e.g., Book Clubs, Goatley, Brock, & Raphael, 1995; Literature Circles, Daniels, 2002). Both types of scaffolding are a part of many discussion approaches (e.g. Instructional Conversations, Thrarp & Gallimore, 1991; Quality Talk, Wilkinson, Soter, & Murphy, 2010) and provide only a few examples of how social constructivist theory has informed classroom discussion research.

**Summary of theories.** In summary, social constructivist theory has ample utility for the study of classroom discussion, as it situates knowledge as the result of collaborative interaction using language as a tool for co-construction of meaning (Almasi & Garas-York, 2009; Palinscar, 1998; Vygotsky, 1978). Constructivist theory is also informative, as it situates knowledge as an internal process of change that takes place when prior schemas are challenged (Piaget, 1932). Although perhaps different with regard to some details, both theories adequately justify discussion as a catalyst for development of higher order critical thinking skills and comprehension. Thus, from either a social constructivist or constructivist perspective, understanding changes in the quantity of student talk is of significant concern (Soter et al., 2008; van Blankenstein et al., 2009). These theoretical frameworks do not, however, provide an adequate model of the mechanisms by which students comprehend what is being read prior to or said in literature discussions. In order to provide a complete theoretical foundation for investigating talk, it is important to review a dominant model of how students comprehend text, via both reading and discourse.
Cognitive Model of Comprehension

Models of comprehension vary due to differences in subject matter or measured outcome (McNamara & Magliano, 2009). For the purposes of my study, I will briefly review the Construction-Integration model (CI). CI is a model of comprehension appropriate to the study of discussion, and is widely accepted (Kintsch, 1988; McNamara & Magliano, 2009; Murphy et al., 2011; Reninger & Wilkinson, 2010). The CI model has been helpful across many empirical studies of comprehension. As McNamara and Magliano (2009) pointed out, however, “it is challenging to describe all of the assumptions germane to a model that has been specified across three books and an uncountable number of experimental tests and demonstrations” (p. 308). Regardless, a few notable examples of empirical support for the underlying assumptions of the CI model will be reviewed next.

Broadly, the CI model is a framework for understanding how a student comes to comprehend text and discourse (Kintsch & Kintsch, 2005; Reninger & Wilkinson, 2010). When a student reads a text the first thing that happens according to the CI model is the immediate recall of that text or related knowledge (i.e., construction). Next, the information is theorized to spread, activating other concepts that are linked to it (i.e., integration) (Kintsch, 1988). The CI model outlines three levels of comprehension, the surface structure, textbase, and the situational model (Kintsch, 1988; McNamara & Magliano, 2009).

The first level, surface structure, is defined as the exact wording and syntax of the discourse or text, which is held in the participant’s memory for a few seconds (McNamara & Magliano, 2009; Murphy et al., 2011). Although surface structure is a critical first step to comprehension, this level is somewhat less important than the other levels of comprehension, as it precludes the creation of meaning (McNamara & Magliano, 2009). Experimental evidence
supports this conceptualization of a surface structure (Clark, 1979; Clark & Sengul, 1979; Jarvella, 1971). Jarvella (1971) studied the processes of immediate recall of long passages of spoken, connected discourse, analyzing the recall of long passages of sentences. It was found that the most recently heard clause (i.e., the immediate clause) was the most likely to be recalled verbatim compared with other sections of the passage. Jarvella (1971) also found that the farther away the phrase, the less likely it was to be recalled with accuracy. This evidence suggested that the surface structure was indeed the initial stage of processing information. Clark and Sengul (1979) found similar results regarding surface structure in their study of pronoun referents. By comparing how readers connected a pronoun to a previously referenced noun, they experimentally showed that the clause was the most basic segment of recalled discourse (Clark & Sengul, 1979). Furthermore, they noted that the final clause read has a “privileged status” (Clark & Sengul, p. 40) compared to other clauses or sentences, meaning it was much easier to connect than other clauses. The results of these studies provided empirical support for the notion that the basic surface structure is the initial level of processing by demonstrating that the construction of information did indeed happen prior to other comprehension mechanisms (Kintsch, 1986).

The second level is referred to as the propositional textbase (Kintsch, 1998; Kintsch & Kintsch, 2005). The propositional textbase is the smallest unit of meaning available to the reader (Kintsch, 1998; McNamara & Magliano, 2009; Murphy et al., 2011). Within the propositional textbase exists both microstructure and macrostructure (Kintsch & van Dijk, 1978). The microstructure is comprised of propositional units linked together to create an interrelated linear sequence of ideas (Kintsch & Kintsch, 2005; Kintsch & van Dijk, 1978). The macrostructure is a hierarchy of the main idea or theme, followed by supporting details. Describing the
macrostructure requires students to synthesize text into a broader informational structure, sometimes referred to as the “gist.” (Kintsch & Kintsch, 2005; McNamara & Magliano, 2009).

The concept of the propositional textbase has been validated by various empirical studies (Haberlandt, Berian, & Sandson, 1980; Kintsch, 1986; Kintsch & van Dijk, 1978; Lesgold, Roth, & Curtis, 1979; Ratcliff & McKoon, 1978). For example, results of experimental research by Haberlandt, Berian, and Sandson (1980) provided support for the CI model’s conceptualization of the propositional textbase at its various levels. The authors investigated how the micro and macrostructures related to each other around boundaries between episodes of discourse (Haberlandt et al., 1980). Results of these experiments showed that readers comprehended episodes as whole units (i.e., macroprocesses), while simultaneously encoding linked, adjacent prepositions in short term memory (i.e., microprocesses) (Haberlandt et al., 1980). Additionally, with regard to microprocesseses of comprehension, results of experiments conducted by Lesgold, Roth, and Curtis (1979) provided additional support for the propositional textbase. Lesgold et al. (1979) investigated the interactions and integration of recently read propositions to surrounding propositions of varying salience. The authors found three ways that integration of a new sentence into comprehension takes place. First, it was found that during reading a small set of directly linked propositions were accessed immediately, second other propositions were understood after a delay when deemed relevant, and finally inference was needed when associated propositions are not directly linked (Lesgold et al., 1979). This process of inference when no salient propositions are available then contributes to the overall difficulty of comprehension and affects recall negatively, as was found in experiments by Kieras (1978). These results provided further support for the CI model construct of the propositional textbase (Kintsch, 1998; Lesgold et al., 1979).
The third level of the CI model of comprehension is the situational model. The situational model is defined as any inference, emotional content, and/or imagery outside of the textbase which is recalled or generated by the reader (Kintsch & Kintsch, 2005; McNamara & Magliano, 2009). A situational model can be created or retrieved when the micro/Macrostructure is integrated with a student’s prior knowledge or emotional state (Kintsch & Kintsch, 2005). The situational model was the last major element of the CI model to be thoroughly researched (Kintsch, 1986). Kintch (1986) empirically investigated the relative contributions of the textbase and situational model in experiments conducted with first grade children. Using arithmetic word problems, Kintch (1986) demonstrated the existence of the situation model, separate from the propositional textbase by examining the errors that children made. Other research has also verified the construct of the “situational model” (Bransford & Franks, 1971; Kintsch, Mandel, & Kozminsky, 1977; Perrig & Kintsch, 1985). For example, Zwaan, Graesser, and Magliano (1995) found that readers monitor both the temporal sequence and causality of the situation model during reading instruction. These findings illustrate how students in discussion may monitor the situation model of a discussion while participating.

**Summary of comprehension model.** The CI model applies to the text discussed and the content of the discussion simultaneously. As comprehension is multidimensional and multifaceted, discussions provide opportunities across various CI levels of comprehension. Student talk production is a dynamic a window into a student’s representation both with regard to the text read and the content of other’s talk. Various discussion approaches map on to this model along a continuum. For example, discussion approaches that emphasize the facts of the text prioritize construction (e.g., Questioning the Author, Beck, McKeown, Hamilton, & Kucan, 1997; Instructional Conversations, Tharpe & Gallimore, 1988). Approaches that emphasize
critical-analytic thinking prioritize integration (e.g., Collaborative Reasoning, Chinn, Anderson, & Waggoner, 2001; Paideia Seminars, National Paideia Center, 2015). Approaches that emphasize personal connections to text may fall between construction and integration (e.g., Book Club, Raphael & McMahon, 1994; Literature Circles, Daniels, 2002; Grand Conversations, Eeds & Wells, 1989). As my study will investigate student discussion, the CI model is a reasonable model for the complex processes that take place during reading and talk about text. Both developmental theories and relevant models of comprehension help define the mechanisms by which students achieve high-level comprehension during discussion. The next section will further focus my review towards research on classroom interactions in order to move towards reviewing research on talk production during small-group literature discussions, specifically.

**Research on Classroom Discussion**

Discussion in the classroom, in its broadest sense, could involve practically any interaction between peers or between teachers and students. It is almost impossible to imagine a school setting in which communication between individuals does not take place. This broad definition does little to guide research on small-group literature discussions, however. Research over the past 40 years has resulted in helpful information regarding classroom talk and classroom researchers have identified various patterns, styles, and habits of discourse (Barnes & Todd, 1977; Cazden, 2001; Mehan, 1982; Shultz et al., 1982; Wells, 1989). The goals of early efforts were to establish what discussions were, how they affected learning, and what types of discussions were the best for particular student outcomes. In the following section, I will discuss current definitions of discussion that have been put forth by leaders in discussion research. Next, major studies that have identified important features of classroom discussion will be reviewed. I will then briefly review research documenting the changing notions of effective discussion from
recitation to dialog. Following this, the research documenting ways to create more discussion based classroom discourse will be reviewed. This will provide the basis for study of outcomes associated with classroom discussions about and around text for development of critical thinking skills.

**Current definitions of discussion.** Current definitions of productive classroom discussion take inspiration from both social constructivist and constructivist theory. Wilkinson (2009) defined discussion as open-ended, collaborative episodes of communication among teachers and students, or among students, for the purpose of fostering student thinking, problem solving, comprehension, or appreciation. Almasi (2002) also defined discussion, calling it “a dialogic classroom event in which students and teachers are cognitively, socially, and affectively engaged in collaboratively constructing meaning or considering alternate interpretations of texts to arrive at new understandings” (p. 420). These two definitions broadly reflect the research on classroom discussion approaches. First, Almasi’s (2002) definition takes into account critical features of engagement and co-creation of meaning. Second, Wilkinson’s (2009) definition adds to this by including student thinking as a goal. These features are broadly agreed upon in every discussion approach that will be reviewed subsequently. Both of these definitions provide an important foundation for what is included in my review and help codify assumptions regarding what constitutes discussion.

**Describing classroom interactions.** These definitions reflect significant research in schools over the past few decades. Early research into classroom discussion identified important patterns of interaction and participation. Mehan (1982) summarized much of the ethnographic research produced prior to the 1980s, establishing foundational concepts that have been used in subsequent studies. He found that more effective classrooms were characterized by students who
have strategies to display the knowledge they have and who know how to interact according to the teacher’s expectations (Mehan, 1982). Mehan (1982) also identified the initiation act, reply act, and evaluation act (IRE) sequence to describe the way the majority of teacher student talk took place in the classroom. This was simultaneously described by Sinclair and Coulthard (1975), who used the terms “initiation, response, and feedback” or IRF interactions. In general, this style of interaction (i.e., IRE/F) was teacher dominated. Specifically, students were observed relying on teacher prompts to speak, and the teacher spoke after almost every student utterance (Cazden, 2001; Mehan, 1982; Sinclair & Coulthard, 1975). Subsequent researchers also described the IRE/F discussion patterns as a “recitation” script, as students were observed simply reciting answers as opposed to engaging in discussion (Langer, 1993; Shachar & Sharan, 1994; Tharp & Gallimore, 1991).

The authors of the ORACLE project made significant contributions to the description of classroom discourse by observing classroom talk on a much larger scale than was typical (Galton, Simon, & Croll, 1980; Mercer & Littleton, 2007). Comprising over three years of systematic observations at 58 primary schools, the authors found that group interactions were often not collaborative. Although students working collaboratively were given an environment that could support the co-construction of knowledge, it was found that they rarely collaborated (Galton et al., 1980). Researchers noted that children worked mostly in parallel on individual tasks instead of collaborating (Galton et al., 1980). Barnes and Todd (1977) added to this research, illuminating typical interactions with a focus on student talk. Significantly, they found that students only engaged in extended, cognitively challenging discussion when outside the apparent control of the teacher (e.g., unstructured time, recess). Students were also found to
have greater ownership over knowledge gained in interaction compared to knowledge presented in a traditional instructional format (Barnes & Todd, 1977).

**Understanding effective classroom interactions.** Early research on classroom interaction was a critical first step towards understanding how students talk and what talk was most productive. Many scholars noted the cognitive impositions on students’ thinking during an IRE/F model interaction (Cazden & Beck, 2003; Nystrand et al., 1993; Tharp & Gallimore, 1988). For example, Mehan (1982) noted that “once students have gained access to the floor, they must know what to do with it” (p. 75). He pointed out that students who do not naturally demonstrate communication skills will be thought of by teachers as “inattentive and unexpressive” (Mehan, 1982, p. 80). Thus, IRE/F interaction was thought to have direct ramifications on how students demonstrate their thinking and how much talk results from classroom interaction. An IRE/F interchange leaves little to discuss, as the answers are predetermined and the teacher controls the discourse (Nystrand, Gamoran, Kachur, & Prendergast, 1997). If opportunities for elaboration are not available, negative consequences are certain to result for more passive students, as well (Almasi & Garas-York, 2009). Classroom discussion that follows IRE/F structure has been shown to be a barrier to learning in that it creates passivity and disengagement (Almasi & Garas-York, 2009; Cazden & Beck, 2003; Murphy et al., 2011). Furthermore, as Nystrand et al. (1997) pointed out, when “recitation starts, remembering and guessing supplant thinking” (p. 6). Tharp and Gallimore (1991) called for classroom talk to move away from IRE/F style interaction so that teachers can adjust dialog to support the development of deeper understanding. In sum, educational scholars widely agree that the IRE/F style of classroom talk runs counter to current educational priorities, which include fostering critical thinking and problem solving (Gillies, 2014; Kuhn, 2015).
**Changing classroom interactions.** Alternatives to the IRE/F interaction were developed as scholars attempted to encourage learning environments where students could benefit from the best of both constructivist and social constructivist theories of learning (Mercer & Littleton, 2007; Michaels et al., 2008). Classroom discourse approaches attempted to structure opportunities for students to both construct their own knowledge (e.g., Tharp & Gallimore, 1991) and benefit from scaffolding by others (e.g., Almasi, 1994; Almasi & Garas-York, 2009; Mercer & Littleton, 2007). An important example of a discussion approach designed to deconstruct the IRE/F model is Accountable Talk (Michaels et al., 2008). Accountable Talk (AT) was based on sociocultural theory, emphasizing the importance of a learning community. One goal of this approach was to change the cultural norms that resulted in IRE/F interaction. It was hypothesized that this would socialize students so that they may hold respectful, grounded discussion and critically evaluate themselves and others. Three strands organize its central features: 1) accountability to the community 2) accountability to knowledge, and 3) accountability to accepted standards of reasoning (Michaels et al., 2008). Accountability to the community involved the direction of discussion towards the group for evaluation, in contrast to simply providing answers for the teacher. This was thought to help students listen to each other and consider others’ perspectives. The strand of accountability to knowledge was defined as how classroom discussion utilized key facts during discussion to support learning “synergistically” (Michaels et al., 2008, p. 291). This concept allowed for students to discuss information even before they fully grasped it. This approach resulted in a strategy for learning simple facts in a socially mediated way. As an example, Michaels et al. (2008) described a student stating 24 was an odd number during an AT discussion. The student with the misconception was then led to the correct idea by other students’ reasoning (Michaels et al. 2008). Finally, the goal of
accountability to reason was thought to encourage students to use logical, well-formulated statements in discussion. To this end, students were guided towards identifying underlying assumptions of others and to critically analyze their own thinking. Additionally, students were encouraged to use logic to justify statements (Michaels et al., 2008). As an example, Michaels et al. (2008) found that kindergarteners using AT could explain reasons and use evidence during a teacher-led discussion about shoe sizes.

The results of research on AT indicated that AT influenced critical thinking in classrooms, although questions remained regarding the effect of this approach on general cognitive ability (Resnick, Michaels, & O’Connor, 2010). Two studies of AT took place in the context of mathematics instruction and science instruction (Resnick et al., 2010). The evidence resulting from these studies suggested that AT student achievement increased on standardized tests of science and math (Adey & Shayer, 2001; Chapin & O’Connor, 2004). Interestingly, results of standardized achievement tests in English also increased, even though AT was implemented in either science or math only (Adey & Shayer, 2001; Chapin & O’Connor, 2004).

There is little available research with regard to AT for literature discussion, specifically. Wolf, Crosson, and Resnick (2005) investigated quality of teacher and student talk during reading comprehension lessons in 21 classrooms across grades one through eight. The goal of this study, however, was to study the relationship between AT and the overall quality of the lesson, defined by the researchers as “academic rigor” (Wolf, Crosson, & Resnick, 2005, p. 31). The results of this study suggested that features of AT were correlated with features of academic rigor, as defined by a researcher developed rubric (Wolf et al., 2005). Participation rates during discussion were not measured, nor were measures of reading comprehension analyzed (Wolf et
Accountable Talk is a broad framework for changing the social norms of discussion and there is little available evidence with regard to reading comprehension or student talk.

The Thinking Together (TT) curriculum is another example of teacher-led, whole class discussion approach. Neil Mercer, Lyn Dawes, Rupert Wegerif, and other scholars developed this curriculum for the purpose of developing critical thinking in classroom interaction in general (Mercer & Littleton, 2007). This approach teaches children to use language as a tool for reasoning, recommends teacher moves to support rich discussion, and encourages an element of “controversy” (Mercer & Littleton, 2007; p. 74) so that students challenge each other. A central goal of the TT approach is to increase “exploratory talk” which was defined as the “embodiment of critical thinking” and “essential for successful participation in ‘educated’ communities of discourse” (Mercer & Littleton, 2007, p. 66). According to Mercer and Littleton (2007) exploratory talk is talk in which participants “engage critically, but constructively with other’s ideas....[in exploratory talk] knowledge is made publically accountable and reasoning is visible” (p. 59).

The results research on TT suggested that TT changed students’ use of language. Specifically, students exposed to TT offered much longer statements and held discussions about issues in much greater detail compared to students in a traditional classroom (Mercer et al., 1999). Research on TT was also carried out on a large sample of nine and ten year-old children using the TT curriculum with science and mathematics (Mercer, Dawes, Wegerif, & Sams, 2004). Results for student talk changes during this study showed significant increases in utterance length for the TT group compared to the control group (Mercer et al., 2004).

These results are relevant to my study of student talk outcomes during literature discussions. TT utilizes peer-peer and teacher-peer interactions that are similar in nature to the
way teachers have led discussions in many literature discussion approaches (e.g., Paideia Seminars, National Paideia Center, 2015; Philosophy for Children, Lipman, 1998). TT encourages students to talk with each other openly and students are taught to consider each other’s ideas. Questioning skills are also an element of the TT curriculum, and students were given support to achieve longer, richer discussions (Mercer et al., 2004).

Although the results of research on TT indicated increases in student talk, the authors did not describe the nature of the increase over time. No study of TT included analysis of change in length of utterance on the individual student level, which may hold important details on who benefited most from the intervention. In addition, it is unclear how the TT curriculum might influence increases in talk differently between students of different reading ability or across gender.

AT and TT were just a few examples of approaches designed to optimize classroom interactions for the development of critical thinking and high-level comprehension. These curriculum interventions were implemented with the goal of changing the way students talk in order to positively influence cognitive development via the cultural tool of language (Mercer & Littleton, 2007; Resnick et al., 2010; Vygotsky, 1978). Results of research on both approaches suggested that student language can be changed and students can be taught to engage with each other in more sophisticated ways. Research that specifically concerns student discussions for reading comprehension will be reviewed next.

**Discussions Designed for Reading Comprehension**

Research and practice broadly support the use of discussion in language arts environments (Gilles & Pierce, 2012; Murphy et al., 2009; National Governors Association Center for Best Practicies & Council of Chief State School Officers, 2010; Nystrand, 2006). The
leadership of the teacher, the cognitive goal, the size of the group, and other factors can differentiate the myriad ways text-based discussions take place (Chinn, Anderson, & Waggoner, 2001; Murphy et al., 2011). For the purposes of my review, I will include discussion approaches that have been the subject of peer-reviewed, empirical research. I will generally follow the inclusion criteria used by Murphy et al. (2009) in their meta-analysis of text-based classroom discussion approaches. As my study concerns student talk outcomes, specifically, studies lacking sufficient detail on student talk were excluded.

First, a brief review of historical context of reading comprehension research will be provided. Following this, I will explain three major categories or stances that can be used to describe text-based discussion approaches. Subsequent sections will be organized by group size utilized for discussion. To begin, studies of whole group literature discussions will be reviewed. Following this section, the research on small-group literature discussions will be reviewed. The primary focus of my review will be on measures of student talk and limitations of the available research. I will show how examining change in talk over time is an appropriate addition to this important literature.

**Historical context.** Beginning in the 1970s, educational researchers sought to uncover the relationships between discussion and reading comprehension. Significant empirical research was conducted to better define reading comprehension and understand its cognitive basis (Nystrand, 2006). Research by the Center for the Study of Reading, the Center for the Improvement of Early Reading Achievement, and the RAND Reading Study group pushed educators to think about reading comprehension in ways far beyond simple recall of facts. The RAND Reading Study Group defined reading comprehension as “the process of simultaneously extracting and constructing meaning through interaction and involvement with written language”
In this context, discussion was recognized as a potential strategy to deepen students’ interactions with text, support high-level comprehension, and foster critical thinking.

**Discussion stances.** The literary stance of a discussion approach towards the texts discussed is a useful parameter for categorizing various approaches (Chinn et al., 2001; Murphy et al., 2009; Rosenblatt, 1994; Soter et al., 2008). In the broadest sense, the stance describes the relationship between the students and the text they are discussing or reading. Rosenblatt (1994) proposed two major stances, efferent and aesthetic. The efferent stance describes reading/discussions designed to simply acquire information (Rosenblatt, 1994). The aesthetic stance described reading with the goal of living through experiences of characters or authors (Rosenblatt, 1994). Chinn et al. (2001) used these two stances to differentiate styles of interaction in their investigation of the Collaborative Reasoning (CR) discussion approach. Chinn et al. (2001) noted another stance, the critical-analytic stance, based in part on research by Wade, Thompson, and Watkins (1994). This stance refers to discussion focused on questioning the text, exploring alternatives, or challenging decisions of characters (Chinn et al., 2001; Wade et al., 1994). In a significant synthesis of the text-based discussion research, Murphy et al. (2009) found the aesthetic stance did not accurately describe any major discussion approaches reviewed. Instead, the authors utilized Jakobson’s (1987) concept of expressive stance. This stance prioritizes a reader’s feelings and emotional connections to literature. The stance of the discussion may affect the nature of the implementation of the discussion. Some approaches prioritize a highly opened structure that results in putting students in control of discussion as much as possible. The Grand Conversations approach (GC) is expressive in nature and emphasizes minimal teacher facilitation. By contrast, the CR discussions approach features more direct teacher involvement perhaps due to the scaffolding necessary to support the critical-
analytic stance. The stances are significant to any review of discussion approaches because they help categorize the goals of the discussion approach and help to explain the context around which measured increases in talk may or may not be observed. Since many approaches have similar broad goals such as supporting comprehension and engagement, it is important to highlight differences among them in underlying assumptions as they relate to how students behave.

Student talk outcomes were analyzed across nine major discussion approaches as a part of a project to synthesize the existing evidence base for text-based discussion (Murphy et al., 2009; Soter et al., 2008). The discussion approaches were categorized by stance and analyzed for indicators of high-level learning and comprehension (Soter et al., 2008). In order to compare approaches, Soter et al. (2008) solicited prototypical transcripts of each discussion approach. Once these transcripts were received, the authors coded each for high-level learning and comprehension indicators. These indicators were determined from extensive reviews of research on discussion features that were most associated with high-level comprehension. In addition, Soter et al. (2008) counted the number of words, turns, and length of utterance (per turn) in transcripts of various discussion approaches. It was found that students contributed the most in approaches with an expressive stance, second most in critical-analytic discussions, and the least in efferent approaches. These findings did not take into account the quality of what was said, however. The length of teacher contribution was also measured. It was also found that while teacher turns were longest during critical-analytic discussions, the total words spoken by teachers and students during these discussions were similar. Soter et al. (2008) hypothesized that these longer turns in critical-analytic discussions resulted from the modeling needed to elicit high-level thinking from students.
Although this research documented the differences in student talk amongst nine major literature discussion approaches, Soter et al. (2008) could not measure student talk production over time. Additionally, the distribution of talk across student participants was not analyzed. A few students may have been responsible for the majority of the discussion, as talk measures were averaged across the groups. The next sections will review relevant research on whole group and small-group literature discussions, focusing on student talk outcomes.

**Whole group text-based discussion.** A significant organizing factor in any review of group discussion is the size of the group of participants (Murphy et al., 2011; Nystrand, 2006; Webb & Palinscar, 1996). Some discussion approaches take place among the entire class, as opposed to within smaller group settings (Murphy et al., 2011; Nystrand, 2006). There is some evidence to suggest that whole group interaction is effective. Van den Branden (2000) compared the comprehension of students who participated in either paired or whole group discussion of a difficult reading passage. The results of this study suggested that greater student comprehension was associated with whole group discussion (Van den Branden, 2000). In addition, Nystrand and Gamoran (1991) conducted large scale research on various elements of classroom interaction. They observed language arts lessons across 58 classrooms in 16 schools and coded for incidents of high quality discussion amongst participants. In this study, greater achievement was found to be associated with classrooms that had higher quality of interaction (i.e., more discussion). The authors also found that students provided more elaborated answers in classrooms where teachers engaged in so called “authentic interactions” (Nystrand & Gamoran, 1991). Nystrand and Gamoran (1991) considered interactions authentic if they were substantively engaging for students and students had input into and control over what was said. It should be noted that this research was designed to explore and observe teacher interaction, not to measure implementation
of any particular group discussion approach (Nystrand & Gamoran, 1991). For the purposes of my study it is relevant to highlight the potential ramifications of whole group language arts discussion approaches on student talk outcomes.

Although a variety of discussion approaches have been piloted in addition to those reviewed here (e.g., Reflective Teaching, Conversational Discussion Groups, Elaborated Interrogation) a comprehensive review is beyond the scope of this project. Next, I will review the research on four specific, whole group literature discussion approaches: Reciprocal Teaching, Questioning the Author, Padiea Seminar, and Philosophy for Children. I will briefly discuss the available research, focusing on details regarding how these approaches affect student talk.

**Reciprocal Teaching.** Reciprocal Teaching (RT), was developed as a comprehensive structure for bolstering comprehension through dialog between a teacher and a group of students (Palinscar & Brown, 1984; Rosenshine & Meister, 1994). RT represents an example of an approach with an efferent stance towards the text, as it is focused on the activation of accurate factual information from the text being discussed. RT occurs between teachers and a group of students about a text. During this dialog, four strategies are emphasized: summarizing, question generating, clarifying, and predicting. An important feature of this discussion approach is the transition from teacher-led instruction to student-student interaction (Rosenshine & Meister, 1994). Students are encouraged to “take over the major thinking role, while the teacher observes and helps only when needed” (Rosenshine & Meister, 1994, p. 481). How long the teacher remains as an active member before students take over may be highly variable in the RT approach. This would have consequences on how much student talk increases are seen over time. In their review of RT, Rosenshine and Meister (1994) reported increases on student reading comprehension as well as increases in the quality of student talk associated with RT. With regard
to comprehension, the students receiving RT were found to be superior to a control group on both standardized and experimenter-developed comprehension measures. With regard to quantity of student talk, only one study assessed student responses. The authors of this study found students to elaborate more, participate more, and ask longer questions after experiencing the RT intervention (Palinscar & Brown, 1984). The increases in responses were only reported qualitatively, and are thus difficult to compare to other measures of student talk in the research literature.

**Questioning The Author.** Another major discussion approach that exemplifies a teacher-led, whole group, efferent discussion approach is called Questioning the Author (QtA). This approach is heavily based on a constructivist model of cognitive development. Its goal is for individuals to make meaning for themselves via discussion techniques (Beck, McKeown, Hamilton, & Kucan, 1997). The developers acknowledge that discussion is not the goal of QtA, simply the “means toward...constructing meaning” (Beck et al., 1997, p. 21). Additionally, QtA discussions take place during reading, as the whole group reads an assigned text. This was designed to directly model the kind of metacognition that students were to engage in when reading independently (Beck et al., 1997). QtA features a significant amount of teacher direction when implemented as developed. Although designed to “engage students with text” (Beck, McKeown, Sandora, Kucan, & Worthy, 1996, p. 391), students in QtA are not as actively encouraged to engage with each other as may be seen in other discussion approaches. For example, during QtA it is appropriate to direct student responses by calling on a participant directly, thus students in QtA do not have as much control over the flow of discussion as in other approaches (Beck, McKeown, Sandora, Kucan, & Worthy, 1996). This may have implications for the amount of student talk that is generated in this approach.
QtA was studied over a year of implementation in two fourth grade classrooms that the authors described as “inner-city” (Beck et al., 1996, p. 385). Evidence was gathered on student and teacher talk outcomes as well as measures of comprehension. A few important effects on student talk were found. During QtA the proportion of teacher talk decreased and the proportion of student talk increased compared to a baseline discussion (Beck et al., 1996). The authors also found that student commentary became more complex in QtA discussions compared to a baseline discussion (Beck et al., 1996). Qualitative data from this investigation indicated increased participation in QtA compared to baseline, even among some students who had experienced little success in school (Beck et al., 1996).

Some other results regarding student talk during QtA have been found. Meta-analysis of research on QtA conducted by Murphy et al. (2009) across available empirical studies found that QtA produced minimal increases in teacher talk and slight decreases in student talk, overall. This may be explained by the dominant role of the teacher in QtA. Soter et al. (2008) also found teacher talk to be greater than student talk in terms of raw words, turns, and length of turn in QtA discussions. Thus, while QtA produces increases in talk, it was found to be a teacher dominated approach (Murphy et al., 2011; Soter et al., 2008). Throughout the available studies of QtA there was limited evidence regarding other explanatory variables for increases in student talk. Furthermore, it was unclear if only a few students were responsible for the average increases in student talk or if student talk increased across many students.

**Paideia Seminars.** Paideia Seminar (PS) is another approach for whole class discussion, implemented as part of the larger Paideia instructional system (National Paideia Center, 2015). Conceived and developed by Dr. Mortimer Adler in the early 1980s, the Paideia instruction system is made up of “intellectual coaching,” “didactic instruction,” and “discussion.” (National
Paideia Center, 2015). PS discussions take a critical-analytic stance towards reading comprehension, in contrast to QtA and RT. One of the major goals of PS is discussion of the ideas and values around the chosen text (Billings & Fitzgerald, 2002). Thus, students create new meanings and connections through active dialog with each other (Billings & Fitzgerald, 2002). The teacher’s role is to promote reasoned criticism between students, as opposed to eliciting answers to questions. Although this is not necessarily a IRE/F style of discussion, the PS method does involve teacher direction that may influence the length of student responses. In addition, most PS discussions have been in high school settings where different dynamics may be at play than in elementary school language arts instruction.

Although many schools around the United States have adopted Paideia, there is somewhat limited peer-reviewed research as to its effects on student outcomes (National Paideia Center, 2015). One notable exception was an investigation of the relationship between PS and writing scores on standardized tests, but the authors did not analyze student talk during PS discussions (Chesser, Gellatly, & Hale, 1997). Only a few studies have analyzed the student talk, specifically. Student talk and qualitative outcomes were analyzed though a case study of PS discussions in an 11th grade honors English class (Billings & Fitzgerald, 2002). Data collected by Billings and Fitzgerald indicated that the teacher talked a greater percentage of time than the students, across the three discussions observed. They also noted that this degree of teacher guidance was higher than recommended by the Paideia trainings (Billings & Fitzgerald, 2002). This was the case even though the participating teacher had been involved in nine days of PS trainings prior to the research study. In addition, Billings and Fitzgerald noted differences in gender with regard to student talk during the Paideia Seminars they observed. According to results of transcript analysis, male students spoke for more turns than female students in two of
the three discussions observed (Billings & Fitzgerald, 2002). Qualitative evidence collected during this study of PS suggested some gender differences in participation as well. For example, the teacher was observed remarking, “I’m tired of hearing talkative guys and quiet girls” (Billings & Fitzgerald, 2002, p. 923). It is unclear if this difference was present in other settings, however, as only one 11th grade classroom case study was investigated (Billings & Fitzgerald, 2002). It is important to note that the age of the students involved in this study may influence the potential effect of gender.

Two unpublished dissertations also collected data on the student talk outcomes of PS discussions (Howard, 1992; Robinson, 2006). In her study of middle school students, Howard (1992) found that PS discussions contained more student talk than a traditional lesson. The author noted, however, that the increase was far from desirable for PS implementation (Howard, 1992). Howard also analyzed student talk associated with both achievement and gender. It was found that PS discussions did not activate participation amongst lower achieving students. During PS discussions lower achieving students made only 15% of spoken remarks in one experimental site and did not participate at all at the other experimental site (Howard, 1992). With regard to talk and gender, the results of Howard’s analysis indicated that males made 81% and 40% of remarks in both sites, respectively.

Robinson (2006) collected talk data on high school students participating in the PS approach for discussion. During the PS discussions studied, she found the teacher took an average of 18% of the talk turns (Robinson, 2006). Additionally, student-to-student interaction was measured and found to comprise 85% of total talk time (Robinson, 2006). Robinson did not provide achievement or demographic data in her analysis. Other meta-analytic research on PS indicated a slight decrease in teacher talk and a slight increase in student talk (Murphy et al.,
2009). Overall, the results of the studies of PS indicate that student talk and participation are somewhat positively affected by PS discussions. Little available evidence exists with regard to how these changes in student talk develop over time and how this change varies. Possible explainers of differences in talk (e.g., gender, reading ability) were noted in a few investigations of this approach. More research is needed to understand changes in student talk over time during classroom discussion.

**Philosophy for Children.** Philosophy for Children (P4C) is an extensive curriculum developed to create “communities of inquiry” (Lipman, 1998, p. 278) based on the Socratic method and social constructivist theory. P4C has a primarily critical-analytic stance towards the texts, which are specific to P4C and designed to instigate an ethical or moral dilemma (Lipman, 1980; Soter et al., 2008). The teacher initiates the discussion by soliciting a topic from the students and asking a broad question (Lipman, 1998; Soter et al., 2008). This approach requires facilitation by the teacher throughout the discussion, and the teacher’s role is not intended to diminish over time. In addition, prior to the discussion students are asked to state their position on the issue in question. This level of teacher control of the discussion may influence the presence or absence of student talk and also may influence changes in quantity of talk over time.

Experimental research on P4C has shown it to be associated with gains in reasoning and reading (Lipman, 1980). To examine effects on reasoning, two groups of 20 fifth-grade students were randomly assigned to either the P4C curriculum, taught by a philosophy teacher, or “business as usual” social studies instruction (Lipman, 1980). After nine weeks of P4C, the treatment group showed statistically significant improvements on a logical reasoning test compared to the control group (Lipman, 1980).
Only one available study of P4C included measures of student talk outcomes, specifically. Chamberlain (1993) studied P4C over 11 weeks with 80 gifted fourth and fifth-grade students, randomly assigned to P4C or a traditional literature group. The treatment and control groups were compared on standardized tests of reasoning (i.e., New Jersey Reasoning Test and Ross Test of Reasoning) and audio tapes of discussions were analyzed (Chamberlain, 1993). The results on the measures of reasoning showed statistically significant increases in the P4C group on the New Jersey Reasoning Test (Chamberlain, 1993). According to transcript analysis, the percentage of student-student talk increased from 15% to 54% of total utterances from baseline to week 11 in the P4C group. By comparison, the control group student-student talk remained constant at 15% of total utterances over the length of the experiment (Chamberlain, 1993). Soter et al.’s (2008) analysis of a prototypical P4C transcript corroborates these results. They found that in the P4C transcript analyzed, students spoke significantly more words than the teacher. Although students spoke more words, the teacher in the sampled P4C transcript took longer turns than the students, on average. Soter et al. (2008) hypothesized this to be due to the critical-analytic stance of P4C, because students may require much more modeling of logic and reasoning skills. In sum, the available data suggest that control is somewhat shared by teachers and student during P4C discussions, even if teachers spoke for longer segments of time (Chamberlain, 1993; Soter et al., 2008). No available research on P4C included evidence regarding student characteristic differences that could explain the increases in talk. Additionally, no studies including analysis of changes in talk over time were found.

**Summary of whole group approaches.** Whole group discussions have been shown to be associated with various positive student outcomes (Applebee, Langer, Nystrand, & Gamoran, 2003; Nystrand & Gamoran, 1991; Van den Branden, 2000). Whole group literature
Interventions differ across their features and result in different kinds of classroom outcomes (Murphy et al., 2011; Nystrand, 2006). Some approaches were associated with increased responses, showing increases in student talk (e.g., P4C, QtA, RT). In other approaches, however, the research reviewed suggested the presence of teacher dominated discussion and some male dominated patterns of discussion (e.g., PS). Regardless, whole group discussions are not expected to disappear from classrooms, thus these approaches represent potential improvements over the IRE/F style of interaction (Nystrand & Gamoran, 1991; Tharp & Gallimore, 1991). Research on literature discussion approaches that took place in small-groups will be reviewed next.

**Small group discussions for reading comprehension.** Early educational research was unclear with regard to the benefits of small-group learning (Nystrand et al., 1993). Some scholars such as Slavin (1980) and Wells (1989) espoused the benefits of small-group learning. Others, by comparison, found no significant benefits in achievement or participation when compared to whole group instruction (Webb & Kenderski, 1984). More recent research has added to the evidence supporting the benefits of small-group interaction. At the current time, the broad consensus is that small-groups are well-suited to literature discussion (Murphy et al., 2011; Phillips & Twardosz, 2003).

Smaller groups have certain advantages over whole group instruction (Linnenbrink-Garcia, Rogat, & Koskey, 2011; Morrow & Smith, 1990; Slavin, 1980; van Blankenstein et al., 2009; Wells, 1989). For example, small-groups put less peer pressure on students than whole groups and make it easier for teachers to praise students (Hudgins & Edelman, 1986; Webb & Palinscar, 1996). Also, small-groups make it impossible for students to shirk responsibility for participation (Webb & Palinscar, 1996). Morrow and Smith (1990) directly compared one-on-
one discussion to small-group discussion with kindergarten and first grade children, finding the small-group condition more beneficial for students’ comprehension of a storybook. Correspondingly, Phillips and Twardosz (2003) found small-group interaction effective in increasing verbal interactions among preschoolers. Sweigart (1991) compared small-group, lecture, and whole group discussion with 58 twelfth-grade students. The small-group condition was found to be associated with much higher quality of discussion and better performance of a subsequent writing task compared to the other conditions. In other research investigating small-groups of high school students, small-group discussions were found to increase performance on simple recall language arts tests (Fall, Webb, & Chudowsky, 2000). With regard to motivation Wu, Anderson, Nguyen-Jahiel, and Miller (2013) found small-group discussions produced greater interest and motivation compared to whole class discussions. The next section of my review will focus on talk outcomes in research on small-group discussion approaches. Both teacher-led and student led discussion group approaches will be reviewed here.

**Grand Conversations.** Eeds and Wells (1989) conducted research on small-group literature discussions with fifth and sixth graders over five weeks. One of their goals was to create a situation in which students would transact with the text and each other. By learning to respond to each other around a text, it was hypothesized that students would become more proficient at interacting during reading (Eeds & Wells, 1989). Termed Grand Conversations (GC), this approach exemplifies an expressive stance towards text (Murphy et al., 2009; Soter et al., 2008). An important feature of this approach is its lack of structure. Although a teacher was present, the teachers were encouraged to let the discussions operate as naturally as possible. Unlike CR or P4C, students were not asked to stake a position before sharing their comments. The authors also chose to forego explicit discussion questions and employed undergraduate
students as facilitators to minimize any notions how of discussions should be conducted (Eeds & Wells, 1989). The goal of the major study of GC was to investigate the ways students talk about stories when explicit teacher guidance was removed and to document the ways students co-constructed meaning beyond literal facts (Eeds & Wells, 1989).

In terms of student talk, the experiment resulted in more student turns than teacher turns in all four small-groups. In three groups the students only spoke slightly more, overall, than the teacher. In one group students produced significantly more talk, and the students spoke 230 more turns than the teacher (Eeds & Wells, 1989). Students also read different stories and had a different teacher in each group. The authors noted potentially influential differences in teacher leadership. Specifically, one teacher was considered much less successful than the others (Eeds & Wells, 1989). Although this study provided evidence that equal student/teacher turns can be achieved in small-groups, it also left some unanswered questions. The authors provided little explanation for why one group was able to achieve more student-led discussion and none of the groups were used as controls for comparison. Additionally, although it took place over five weeks, talk data was not reported for each time point. This leaves unanswered the question of how student talk developed over time in this approach. Analysis of a GC transcript by Soter et al. (2008) was in agreement with the results of the research reviewed here, finding more student talk than teacher talk in the GC approach.

**Book Club.** The Book Club (BC) approach is an example of an expressive approach to literature discussion (Murphy et al., 2009; Raphael & McMahon, 1994; Soter et al., 2008). BC discussions comprise one element of a larger language arts program that includes writing, instruction, and whole-class discussion. These elements all are intended to support students’ small-group discussions (Goatley, Brock, & Raphael, 1995; Raphael & McMahon, 1994). BC
was based on social constructivist models of learning. To this end one goal of BC is the creation of a “community of readers” (Raphael & McMahon, 1994; p. 103) rather than emphasizing individuals. Importantly, BC discussions are entirely student-led, instead of facilitated by a teacher (Raphael & McMahon, 1994). This may impact the amount of talk that results, of course, because students in BC experience less teacher scaffolding during discussion.

Goatley et al. (1995) studied a five member group of diverse learners participating in BC during the final three weeks of the school year. The authors conducted frequency analysis of turns, by participant. The data showed that the two group members with the most experience in BC discussions contributed the most. Specifically, in four out of the five discussions, the same two experienced BC members spoke over 50% of the total turns (Goatley et al., 1995). The authors also noted that the two dominant students displayed leadership in different ways. One student had more background knowledge, while the other had pro-social skills that were frequently used to re-direct the conversation back to the topic (Goatley et al., 1995). Some change in interaction was qualitatively documented in this study in one case. Specifically, a relatively more aggressive student with social skill deficits changed his style of interaction to more appropriate methods during the three weeks of discussion (Goatley et al., 1995). It is notable as well that the contributions of the two more quiet students in the group did not change over time, according to this study (Goatley et al., 1995). Although the authors noted that quiet students often made appropriate contributions, these findings contrast the assumption that discussions result in more equal participation. Although all the group members engaged appropriately, this study also raises questions about how discussions can alleviate disparities in participation due to students’ prior discussion experiences (Goatley et al., 1995).
**Literature Circles.** Literature Circles (LC) is an expressive approach that shares many of the same traits as the Book Club approach. This approach emphasizes the creation of authentic discussions and engaged habits of reading amongst participants (Daniels, 2002). Importantly, however, this approach includes the assignment of roles to students to stimulate discussion. Some examples include “passage picker,” “connector,” and “question asker” (Daniels, 2002, p. 103). Many studies of LC reported increases in engagement and positive changes in reading comprehension associated with this approach (Carrison & Ernst-Slavit, 2005; Davis, Resta, Davis, & Camacho, 2001; Manning, 2010).

LC has been studied quantitatively to a greater degree than other expressive approaches, and some of that research includes student talk data. Murphy et al. (2009) found LC had strong positive effects on student talk in their meta-analysis of research on discussion approaches. Other research has produced evidence that students’ oral communication abilities were bolstered by LC discussions (Davis et al., 2001; Farinacci, 1998). With regard to student talk data, specifically, Soter et al. (2008) found that students spoke more and took more turns than the teacher (Soter et al., 2008). The length of student utterance was similar to some other discussion approaches and did not differ from the length of teacher utterance (Soter et al., 2008).

Although the data on talk outcomes during LC discussions is promising, LC does not compare to other small-group discussion approaches for measuring student talk. Due to the fact that student roles are assigned, student talk contribution will certainly be affected by these roles. Thus, student talk data from LC research, while important, is not as relevant as data from approaches wherein all students occupy the same role in shared discussion. Generally, however, LC is another example of a discussion approach that increases student talk, even if attenuated by a role assigned for discussion.
Instructional Conversations. Tharp and Gallimore (1988) proposed a model of interaction called Instructional Conversations (IC). The IC model includes guidance for classroom discussion in terms of instructional and conversational elements. Some of the instructional elements include language stipulating that the instructor should help students use evidence to support their positions and elaborate their statements. The conversational elements encourage “open-ended questions,” a “challenging atmosphere,” and “self-selected turns” (Goldenberg, 1993, p. 319). IC is an example of a discussion approach with a primarily efferent stance, as its instructional component is heavily based on constructivist models of knowledge building. The conversational elements have a somewhat critical-analytic stance as well, however, some studies implementing IC appeared to prioritize efferent goals (Echevarria, 1995; Saunders & Goldenberg, 1999, 2007; Soter et al., 2008). IC allows students more control over the flow of discussion than some other critical-analytic approaches like CR or P4C. The teacher is still present in the discussion groups, however, so as to provide support when needed. This may have ramifications for the amount of talk resulting from IC.

Though the original IC guidelines did not include a strict prescription for small-group implementation, many studies of IC utilize a small-group structure (Tharp & Gallimore, 1991). Saunders and Goldenberg (1999, 2007) studied the use of IC to bolster reading comprehension among small-groups of Latino students who were still emergent English speakers. Echevarria’s (1995) research on IC was conducted in a self-contained special education classroom with only five students. One other study of IC by Saunders and Goldenberg (1999) was implemented using the teachers’ pre-existing heterogeneous small-group rotation system. Another important consideration is that IC was studied primarily with Latino/Hispanic students and Latino/Hispanic

Overall, the results of the research on IC suggested that IC had positive effects on general comprehension and text explicit comprehension (Echevarria, 1995; Murphy et al., 2009; Saunders & Goldenberg, 2007). Analysis of student talk outcomes was included in a few of these studies. Saunders and Goldenberg (2007) conducted a study of IC with one class of fourth-grade English Language Learners in order to compare IC to traditional literature instruction. Students were randomly assigned to receive either IC instruction or IRE/F style discussion. Results of this experiment indicated that the IC students’ average utterances were almost as long as the teacher’s utterances (Saunders & Goldenberg, 2007). The students’ utterances during IRE/F style discussion were much shorter, only a third as long as the teacher’s (Saunders & Goldenberg, 2007). This is impressive considering the assumed fluency difficulties these students experienced as English language learners (ELL). Further supporting these findings, research on IC was conducted in a special education setting among students identified as having learning disabilities (Echevarria, 1995). Student utterances were measured and positive increases were found for IC lessons compared to a traditional literature lessons. Specifically, Echevarria (1995) found a higher number of utterances in IC and that the utterances were longer. Qualitative data from the research on IC indicated that participation during IC increased and students contributed more without teacher prompting (Echevarria, 1995; Saunders & Goldenberg, 2007). Meta-analysis of research on IC found increases in student talk and decreased teacher talk as well (Murphy et al., 2009).

This is significant, however, typical transcript data analyzed by Soter et al. (2008) provided another perspective on IC. They found that teacher utterances were longer than student
utterances, and that the teacher spoke more words than the students. Considering, however, the ability levels of the students participating in IC research, it was hypothesized that teachers may have needed to model, scaffold, and guide discussion during IC. Overall, the available evidence suggested that using IC for literature instruction results in increased student talk, though possibly not in comparison to teacher talk. The student talk outcomes in these studies were not analyzed over a significant length of time, however, so it is unclear if they were stable. It is also unclear if only a few students accounted for a significant proportion of talk produced, or if hesitant students made increases in talk due to participating in IC.

**Collaborative Reasoning.** Collaborative Reasoning (CR) is a teacher facilitated small-group discussion approach with critical-analytic stance towards the text (Chinn et al., 2001). The developers of CR were inspired in part by research on discussion of important historical issues among adults (see Wade, Thompson, & Watkins, 1994). CR was conceived as a discussion approach that encourages children to focus on a major dilemma, consider the characters’ reasoning for choices, and appeal to evidence to support positions (Chinn et al., 2001). Open participation is encouraged in CR discussions because one goal of this approach is to develop “values and habits of mind...for choosing among competing ideas” (Anderson, Chinn, Waggoner, & Nguyen, 1998, p. 172). Teachers lead these discussions and assist students in taking a position that a group will collaboratively consider via the CR discussion (Waggoner, Chinn, Yi, & Anderson, 1995). The teachers in CR maintain a certain amount of control over turns, at least to initiate discussion. Due to the fact that students are asked to state positions at the outset, the students may not have as much control over either the flow of discussion or the negotiation of positions different than what they stated at the beginning of each CR discussion.
CR has been studied across a variety of outcomes, including measures of argumentation, measures of transfer to written argument, and measures of motivation (Chinn et al., 2001; Reznitskaya et al., 2001; Wu et al., 2013). One study by Chinn et al. (2001) conducted detailed analysis of effects on student talk associated with CR discussions. This study took place over a seven week period, comprising a total of 12 discussions across 10 classes of fourth-grade students (Chinn et al., 2001). The discussions were videotaped and coded for turns and also the rate of speaking was calculated by dividing total words spoken by total time in discussion (Chinn et al., 2001). Few studies of discussion have utilized this many waves of data collection.

Interjections were also measured, as determined by a rubric developed by the authors comprised of “back-channeling,” “failed attempts to gain the floor without interrupting,” “interruptions,” and “interjected comments” (Chinn et al., 2001, p. 393). The authors measured turn length as well to understand if students spoke more per utterance when participating in CR.

Results of this study of CR suggested that increased student talk, decreased teacher talk, and that students’ utterances followed each other more frequently compared to baseline (i.e., students spoke back to back without a teacher interjection). With regard to student utterances, the authors found statistically significant increases in student talk compared to a baseline discussion, increasing from 66 words per minute to 111 words per minute (Chinn et al., 2001). Also, the percentage of turns taken by teachers fell by 13% during the CR approach. Runs of student turns increased as well, indicating a marked shift away from IRE/F style discussion. Importantly, Chinn et al. (2001) sought to understand change in student talk on the individual student level. Their findings revealed variability in how talk increased amongst the students. Although 46.7% of students made modest to high increases, 52.3% students either talked the same amount or decreased their talk (Chinn et al., 2001). The authors did not find any differences between
genders in talk outcomes (Chinn et al., 2001). This finding contrasts some held assumptions regarding gendered talk in classroom talk and the findings of other studies reviewed here (Billings & Fitzgerald, 2002; E. Cohen & Lotan, 1995; Hammersley, 1990).

Finally, when analyzing the distribution of length of utterance, the authors found no differences in the length of utterance between baseline discussions and the CR discussions (Chinn et al., 2001). This is surprising, considering some research has suggested that students contribute longer utterances when given more control (Nystrand et al., 1997; Webb, 1991). By contrast, the CR discussion approach did not result in increased mean length of utterance compared to traditional discussion, even though students did speak more overall (Chinn et al., 2001). This may be a result of the way CR discussions operate, since individual students are proving previously established positions. A resulting lack of increase in mean length of utterance may be due to the fact that students in CR were proving their positions instead of elaborating on possible new insights through the group discussion.

In the analysis of an exemplary CR transcript conducted by Soter et al. (2008), they found that students took more turns than the teacher, but that teacher average length of turn was longer. It was hypothesized that critical analytic approaches require levels of teacher modeling that result in longer teacher turns (Soter et al., 2008). In sum, the CR discussion approach appeared to increase student talk, overall, compared to a traditional discussion condition. Some data showed that the length of student utterance, on average, did not increase during CR discussions (Chinn et al., 2001). It is possible, however, that the treatment condition during this study of CR was not long enough to produce effects on length of student talk. In addition, students took a position in CR and then supported it, not necessarily co-constructing their initial positions amongst each other.
Summary of small-group approaches. In conclusion, the results of relevant research indicate that small-group discussion approaches are associated with increases in student talk. In many studies, the basic measures of student talk appear to increase with participation in small-group literature discussion. There is, however, variability in relative amount of teacher talk and variability in how the increases happened. In some cases, students talked more, but their utterances did not change in length. This may indicate that they are simply communicating efficiently. It is also possible that students were not giving extended explanations, even when given the opportunity. This is of concern, as social constructivist theory would suggest that longer utterances are desirable for students to co-create meaning from texts (Almasi & Garas-York, 2009; Eeds & Wells, 1989; van Blankenstein et al., 2009; Webb, 1991). In addition, some of the results of research reviewed suggested a few students were responsible for overall average increases in student talk. This would indicate that the equitable engagement in small-group discussion did not happen. It might be that longer participation in the small-group is necessary for children to equalize their contributions. Due in part to the limitations noted here, my study of change in student talk over time represents an important addition the literature. The span of measurement occasions for my study also sets it apart from research reviewed here, as discussions were measured over almost an entire school year. My study will model the trajectory of talk increases in a small-group literature discussion approach and investigate possible explanations of variation in these trajectories. Furthermore, few scholars have investigated how student talk varies by other factors that may influence talk. In the next section, I will review the key features of Quality Talk, a small-group literature discussion approach developed from research on classroom discussions. Following this section, I will review the relevant research on factors that influence how participation in small-groups occurs.
Quality Talk

Quality Talk is a model of small-group literature discussion designed to promote high-level comprehension through talk about and around texts (Wilkinson et al., 2010). It was created using data gathered from a landmark meta-analysis of existing discussion approaches (Murphy et al., 2009; Wilkinson et al., 2010). One critical finding of this meta-analysis was that while most discussion approaches resulted in increased student talk, only a particular type of talk was associated with high-level comprehension (Murphy et al., 2009; Wilkinson et al., 2010). Few approaches were found to increase both basic and high-level comprehension, and some were found to be effective at increasing critical-thinking (Murphy et al., 2009). The development of Quality Talk was an attempt to systematically create the type of discussion shown to be most effective by research. The creators of Quality Talk considered the features of the most effective approaches such as utilizing a small-group structure, a critical-analytic stance towards text, and an active teacher role (Wilkinson et al., 2010). Quality Talk is best thought of as the combination of four key elements. These are the ideal instructional frame for discussion, specific pedagogical principles based on research, teacher moves found to be effective, and empirically supported discourse tools (Wilkinson et al., 2010).

The instructional frame refers to the conditions necessary for a high quality classroom discussion to take place. One of these conditions is shared control over discussions. In Quality Talk, the teacher maintains control over choice of text that is to be discussed. Students in Quality Talk balance the power of the teacher by exercising control over the interpretation of the text and the turns taken (i.e., the flow of discussion). In addition, rather than being exclusively critical-analytic in terms of stance, Quality Talk does not restrict student responses that may be
expressive or efferent in nature. These responses may serve to foster higher level comprehension by providing important avenues for engagement with texts (Wilkinson et al., 2010).

Certain pedagogical principles are requisite to generating a culture of inquiry and exploration, and these principles are another critical feature of the Quality Talk model. In order to create a space were students actively co-create meaning through discussion, a culture has to be established which encourages interaction, engagement, and mutual respect. Examples of these principles include the use of mutually agreed upon ground rules for discussion, utilizing interesting texts, and asking broad, engaging questions. The pedagogical principles include teacher and student attitudes towards talk. Namely, talk is a tool for thinking and students are encouraged to co-create meaning in discussion through challenging each other.

To ensure that these discussions are productive, Quality Talk includes explicit teacher moves. Teacher moves refer to how discussion leaders (i.e., teachers) facilitate productive talk in discussion. As students require modeling and scaffolding to become skilled discussion participants, teachers must judiciously employ strategies such as summarizing, modeling, and challenging. In this way Quality Talk teachers model relating to the text so that students can begin to internalize these skills for themselves. The developers of Quality Talk found that teachers benefited from the identification of explicit techniques for supporting productive talk (Wilkinson et al., 2010).

The discourse elements emphasized by Quality Talk were culled from extensive reviews of literature on discussion approaches to comprehension of text. In addition to a major meta-analysis, the developers of Quality Talk analyzed discussion approaches for outcomes shown to be linked to high-level comprehension (Wilkinson et al., 2010). Some of these include asking authentic questions, building on another student’s statement (i.e., uptake), and elaboration of an
idea (Mercer et al., 1999; Nystrand et al., 2003). These elements are taught to students via whole group lessons, and then practiced in the small-group context. Teachers in Quality Talk classrooms are provided with professional development in discourse elements to effectively support discussion.

Overall, these four elements define an environment in which students can experience the cognitive benefits of discussion. These elements range from simple classroom behaviors to complex attitudes towards knowledge. Importantly, the developers of Quality Talk utilized research to create a model using elements identified across decades of research. The Quality Talk model is a codification and implementation of these elements and an appropriate context for studying the development of student talk over time in small-group literature discussions. During any small-group interaction, variables other than the features of the approach may explain differences in how much students speak.

Factors That Influence Student Talk

A variety of factors may influence student talk in small-group literature discussions and participation in collaborative learning, generally (Webb & Palinscar, 1996). Small-group outcomes are shaped by a combination of individual and group characteristics (Webb & Palinscar, 1996). Small-group participation theory includes useful ideas for contextualizing the talk that students produce in small-group discussions. Variables such as member attributes, group characteristics, status, culture, group size, and task may affect participation (Bonito & Hollingshead, 1997). This review will focus on only a few of these, specifically, status and text type (i.e., task).

Variables that may generate insight into differences in talk outcomes fall into two major categories: status and task. Two relevant factors fall within status, specifically, gender and ability
level. Educational research has frequently included gender as a variable that can influence outcomes (Bonito & Hollingshead, 1997; Cazden, 2001; E. Cohen, 1984; Webb & Palinscar, 1996). Reading ability is also an important variable to consider and has been shown to have bearing on small-group outcomes (Webb & Palinscar, 1996). Task is also an important factor in participation (i.e., student talk outcomes) (Bonito & Hollingshead, 1997). Although the broad task during Quality Talk (i.e., discussion about and around text) did not change over time, the type of text being discussed varied from week to week. Due to the fact that different types of text are read for different purposes, the task can be thought of as changing along with text type for the purposes of my study (National Assessment Governing Board, 2012). Furthermore, results of research have suggested that variation in student talk outcomes is associated with the type of text (Leal, 1992; Li et al., 2014).

I will first briefly review the theory of how these characteristics influence students, generally, so as to better situate the variables that I examined. Next, with regard to status, the research around reading ability as it affects student talk will be reviewed. Subsequently, I will review the research on gender as a status characteristic that may affect student talk. Finally, with regard to task, the research around text type as it relates to talk outcomes in small-group discussions will be reviewed.

**Expectation states theory.** Expectation states theory (EST) is a general sociological theory that readily helps to explain certain phenomena occurring during classroom interaction (Berger, Rosenholtz, & Zelditch, 1980; E. Cohen, 1984). Especially in small-group interaction, EST helps to explain patterns of dominance or heightened participation in group interaction even when status differences are irrelevant to the task at hand (Berger et al., 1980; E. Cohen & Lotan, 1995). Status in EST is comprised of characteristics that influence the way one is perceived
relative to the social context. A few examples of status characteristics that may influence interaction include, but are not limited to, gender, race, ability, education, or physical attractiveness (Berger et al., 1980). Furthermore, status characteristics may be “specific” or “diffuse” (Berger et al., 1980, p. 482). Using EST to analyze classroom interactions, Cohen and Lotan (1995) presented “reading ability” (p. 101) as an example of a specific status characteristic. In their research, they found that reading proficiency directly influenced an expectation of proficiency among group members. Diffuse characteristics, by contrast, require a generalization based on social norms and assumptions. An example of a diffuse status characteristic is gender, since the status expectations are based on group members’ generalizations regarding all members of that gender (Berger et al., 1980). In diffuse status interactions it is thought that expectations of competence or incompetence are just as active as in specific status interactions, even though these expectations are founded on socially constructed assumptions (Berger et al., 1980; E. Cohen & Lotan, 1995). Thus, based on EST, it is assumed that students have expectations of competence during small-group reading discussions that will influence their talk. The two status characteristics that I will investigate are reading ability and gender.

Reading ability. As mentioned, reading ability is a specific status characteristic that can affect interaction. Cohen (1984) noted that reading ability is “of central importance in elementary school classrooms because it often becomes a prerequisite for successful participation...” (p. 173). Current research on student talk output is unclear with regard to how higher or lower reading ability students interact. small-group interaction researchers systematically investigated the interaction of academic ability with relation to the group and amount of explanation (Webb & Kenderski, 1984). Webb and Kenderski (1984) found that relative academic ability was
positively related to longer utterances, specifically, giving explanations for ideas (Webb & Kenderski, 1984). This association may persist even when ability is not necessarily authentic. For example, a study by Dembo and McAuliffe (1987) used fake academic tests to manipulate the perceptions of ability amongst a group of students. They found that the students who were perceived by the group as higher skilled dominated small-group tasks, even if their actual ability was lower (Dembo & McAuliffe, 1987). Ability differences are relevant to talk outcomes because reading ability and general academic ability are powerful forces in classroom interactions (E. Cohen, 1984; E. Cohen & Lotan, 1995; Dembo & McAuliffe, 1987).

Scholars specific to small-group literature discussions have yet to fully investigate how differences in ability may affect talk outcomes across discussion approaches. The measures of ability that have been used in prior research are somewhat vague and often subjective. Few studies on discussion approaches have included a standardized ability measure such as oral reading fluency. Of the evidence that does exist, little of it is on the individual student level. In one case, students were simply dichotomized to either a “high” category or a “low” category (Chinn et al., 2001). Chinn et al. (2001) found that the low achieving group was no different from the high achieving group with regard to increases in talk (Chinn et al., 2001). Due to the fact that CR discussions utilized a relatively high level of teacher control, however, variation due to ability may not have been observable during implementation of this discussion approach.

In their study of IC, Saunders and Goldenberg (2007) categorized ELL students into high, medium, and low language ability groups. They found that English proficiency did not predict talk output in IC discussions (Saunders & Goldenberg, 2007). The ratio of more participatory to less participatory students stayed constant across the three ability groups, though all of these students were in some way less proficient than native speakers (Saunders & Goldenberg, 2007).
Importantly, the cultural makeup of these groups was somewhat homogeneous, as the study was conducted in a 93% Hispanic school. Their study did not include analysis of individual student reading ability based on normative measures, as levels were assigned based on teachers’ judgments of proficiency (Saunders & Goldenberg, 2007).

When analyzing BC discussions with a small-group of diverse elementary level students, Goatley et al. (1995) found mixed results with regard to general achievement as it related to participation. Of the two most dominant students, one had relatively good reading skills, while the other was reported to struggle with comprehension (Goatley et al., 1995). Although these students varied in the nature of their participation, they accounted for the majority of turns taken. Notably, the student in this study with the most academic difficulty participated almost as much as the two most dominant students (Goatley et al., 1995). One participant in BC discussions with average reading ability was observed participating the least of the five students (Goatley et al., 1995). This research does not show how differences in ability may have influenced talk output over time, however. Also, no standardized measures of reading ability were reported in this research.

In contrast, Howard (1992) found that lower achieving middle school students did not participate in PS discussion at all. The measure used to categorize achievement in this study was the California Achievement Test. Two groups of middle school students were studied while participating in PS. In one group, only 15% of total spoken remarks were made by low achieving students and in the other group, low achieving students made no contributions at all (Howard, 1992). As the PS discussions were conducted in a whole group setting (i.e., approximately 20 students), it may be that it was simply easier for low achieving students to avoid participation. It is not known whether these students would have participated more in a small-group setting.
Furthermore, only one occasion of a PS discussion was measured in Howard’s study, so it is unclear whether the lower achieving students would begin to participate more over time.

In sum, there is limited evidence in the current research regarding associations between standardized measures of reading ability and student talk output. It is possible that the age of the students may change the intensity of the influence of ability. During discussions with elementary age students, few differences in talk associated with ability were were found. This was the case in approaches that are critical-analytic (CR) and efferent (BC, IC). In Howard’s study of PS discussions with middle school age students, however, ability appeared to have negative associations with talk. All in all, although some data has suggested that ability may have no effect on how talkative a student is, the existing literature does not fully examine how a discussion approach may change a student’s talk output over time with respect to a student’s initial reading proficiency. Next, I will review the empirical research with regard to the influence of gender on talk output in small-group discussions.

**Gender.** Gender has been widely studied with regard to its influence on interactions in the classroom. Some evidence has been found that gender may influence participation in group interaction (Webb & Palinscar, 1996). Gender can be thought of as a diffuse status characteristic, meaning that it can affect status expectations even though it is not directly relevant to the task (Berger et al., 1980). These expectations are powerful and pervasive. As Canaan (1990) noted, classroom power disparities reflect the dominant U.S. societal norms, wherein men are “public” while women have a “domestic” (p. 217) orientation. This may result in expected initiative amongst boys, while girls are expected to contribute less (Canaan, 1990). Swann and Graddol (1988) wrote that this difference results from a “complex social process which seems to endow men with greater power than women in social interaction” (p. 153).
With regard to whole group traditional classroom interaction, often boys have been found to dominate in a variety of ways (Drudy & Chathain, 2002; Howe, 1997; Swann & Graddol, 1988). These ways include calling out answers, misbehavior, and raising their hands more frequently (Drudy & Chathain, 2002; Howe, 1997; Swann & Graddol, 1988). These results were similar to those found among 10 and 11 year olds by French and French (1984). Many of these studies concluded that these phenomena resulted in discouraging girls from classroom discourse, thus robbing them of critical chances to develop initiative and to demonstrate higher-level thinking skills. Howe (1997) pointed out that this disenfranchisement may negatively influence attitudes towards learning and negatively bias girls’ participation in discussions as they become adults. The research is somewhat inconclusive, however, as Hammersley (1990) called into question two of these studies (e.g., Swann & Graddol, 1988; French & French, 1984).

Hammersley (1990) pointed out that the overall average imbalance of participation noted by French and French (1984) was due to a small percentage of boys who showed exaggerated attention-seeking behaviors. In the data collected by French and French (1984), this small-group of boys gained attention/participated more than the other boys as well as girls in the classroom. When this small-group of boys was taken out of the analysis, the gender imbalance became far less dramatic (Hammersley, 1990). In addition, French and French (1984) only measured the number of initiations of talk, not the length of turn taken. This means that even if some students spoke for a longer time (i.e., participating more) this extended contribution was not captured by their analysis (French & French, 1984; Hammersley, 1990). A similar study of elementary school students conducted by Swann and Graddol (1988), improved on the work of French and French (1984) by measuring duration of talk during the lesson observed. Their study was conducted on only one lesson, however, so it is difficult to generalize their findings to other classroom
interactions (French & French, 1984). These studies were conducted in a whole group setting, but the dynamics of gender in small-group interaction have been studied as well.

The research on gender differences in small-group work appears to somewhat follow the patterns found with whole group classroom interactions (Webb & Palinscar, 1996). In examining collaborative groups, Lockheed and Harris (1984) analyzed the benefits of participation in mixed-sex groups compared with same sex groups in fourth and fifth grade students. They found that mixed sex small-groups benefited males’ participation and dampened females’ participation (Lockheed & Harris, 1984). In addition, Fisher (1994) investigated a small-group activity amongst high school students and found boys to talk more than girls. This study, unfortunately, was a case study with a very small sample size. Contrasting results regarding small-group work have been found in other research. Specifically, Webb and Kenderski (1985) found no gender differences in small-group participation amongst low achieving African-American elementary school students.

Although many investigations of gendered differences in talk exist in the research literature, there have been relatively few studies of how literature discussion approaches may change these dynamics. Of the approaches for small-group literature discussion reviewed here, few note any differences in gender. For example, no sex differences in talk were found during the implementation of CR discussions amongst fourth grade students (Chinn et al., 2001). It must be noted that CR discussions were more controlled by the teacher than some other discussion approaches. This may have masked the influence of gender on interaction. By contrast, in a qualitative case study of a fifth grade classroom, Evans, Alvermann, and Anders (1997) found small-group literature discussions to be a site of significant disempowerment and silencing of girl participants. These findings add to the research documenting the dominance of boys in
mixed sex groups, and casts some doubt on how well literature discussions provide an equitable environment (Evans et al., 1997). Evans et al. did not report gender differences via quantitative measures. In discussions in the whole group setting, Billings and Fitzgerald (2002) noted qualitative evidence to suggest that high school boys were more dominant during PS discussions. Results of Howard’s (1992) research on PS discussion in the middle school setting indicated that males made the majority of comments in one site studied, but less than half of remarks in the other. Overall, the differences in participation in small-group literature discussions between genders are unclear. Gender is a status characteristic that has been shown to be important, but the specifics of its influence on small-group talk outcomes are uncertain. In addition, some evidence has suggested that the interaction of the discussion approach and the age of the participants could influence the potential effect of gender on participation. Even so, in the research reviewed here no studies examined how gender may predict change in student talk over time.

**Text type.** Research into the effects of the text type on talk outcomes in small-group literature discussion is limited when compared to the research on the effect of text type on comprehension (Li et al., 2014). Leal (1992) studied differences in discussions of three types of text, finding informational-narrative hybrids to have more potential for discussion than simple narrative or informational texts. This was a small, cross-sectional study, however, using three experimental texts with small-groups of fifth grade students (Leal, 1992). As it was not a counterbalanced design, the choice of these texts may have had other effects on group discussion due to differences in student interest (Leal, 1992). Text type can change how teachers lead discussions, as well (Price, Bradley, & Smith, 2012).

Li et al. (2014) examined teacher talk and student talk for indicators of high-level comprehension during fourth/fifth grade small-group literature discussions that took place over a
year. The discussion approach used was Quality Talk, making this research very relevant to my study. Discussions took place around reading selections from the classroom’s basal reader series, and texts were identified as either narrative or informational in structure (Li et al., 2014). These determinations were made based on separation of text types utilized in the NAEP assessment framework (National Assessment Governing Board, 2012). As articulated by the authors of the NAEP, the text types differ in terms of both structure and purpose. With regard to structure, narrative text has structures (i.e., story grammar) which sets it apart from informational texts. Informational texts are usually more organized and hierarchical (National Assessment Governing Board, 2012). In addition, the purpose differentiates these text types. Narrative texts may be read for pleasure or for new perspectives, while informational texts are read to gain knowledge (National Assessment Governing Board, 2012). The analysis conducted by Li et al. (2014) revealed that students’ talk contained more markers of high-level comprehension during narrative texts and teachers asked more authentic questions about informational texts. Also relevant to my study were their findings with regard to “elaborated explanations.” Elaborated explanations were defined as longer segments of student talk where a single student provided multiple statements about a topic (Li et al., 2014; Webb, 1991). Elaborated explanations have been determined to be a marker of high-level comprehension and thus desirable for small-group literature discussion (Soter et al., 2008; Webb, 1991). Li et al. (2014) found the number of elaborated explanations to be statistically significantly higher in discussions of narrative texts. The authors noted that during narrative discussions students may have related personal experiences more readily, resulting in longer instances of student talk (Li et al., 2014). By contrast, informational texts elicited student talk and explanations around facts. Since facts from the story were being negotiated and clarified via discussion, this may have resulted in shorter talk.
outcomes (Li et al., 2014). This research is significant to my study as it clearly captures differences in student talk associated with text type. Some limitations of this research exist, however. For example, this study did not examine changes in student or teacher talk over time. In addition, the authors only coded the middle 10 minutes of each discussion to ensure consistency of analysis (Li et al., 2014). While this makes sense, important student talk may have been lost using this approach. In summary, although limited research has investigated the relationship between text type and small-group literature discussions, it stands to reason that responses to various texts will affect how much students speak about those texts. The types of text that have been analyzed can be broadly thought of as falling into either the narrative or informational, both of which may influence student talk.

**Summary of factors that influence student talk.** Student talk outcomes and group participation may be associated with various factors that originate both with students and from the texts they read. Group participation (i.e., student talk) may be understood as a function of Expectation States theory, which frames participation in terms of status. Based on EST, status characteristics such as gender and ability have been identified as critical factors that may explain how much students talk. In addition, type of text may influence how much students talk in small-group discussions. These areas are relevant to my study and will be investigated to determine any possible impact on student talk.

**Summary of Relevant Literature**

Relevant research about and around small-group literature discussions was reviewed here in order to provide a contextual foundation for the study of talk outcomes associated with the Quality Talk approach to discussion. Broadly speaking, most empirical research has shown that discussion approaches tend to be associated with increases in student talk. These increases may
be different based on the stance of the discussion and the size of the group. Indicators of high-level comprehension were found to vary by discussion approach as well, however, almost no studies have investigated change in student talk at the individual level (Soter et al., 2008). Researchers have sought to optimize how children discuss texts as educational priorities shifted towards more sophisticated comprehension goals. Many discussion approaches for reading comprehension outcomes have been studied and discussion continues to enjoy broad popularity with educators looking to bolster high-level comprehension (Gillies, 2014).

Research on interventions to generally support critical thinking in classroom interaction was reviewed as well. The reviewed interventions demonstrated that students’ talk may become more rich and explanatory when supported with thoughtful changes in the expectations placed on small-groups (i.e., Thinking Together, Accountable Talk) (Mercer & Littleton, 2007; Michaels et al., 2008; Resnick et al., 2010). None of the research or innovation in classroom discussion would exist without robust theoretical foundations. Constructivist theory provides a basis to understand individual mechanisms for learning and development in interaction. Social constructivist theories of learning provide frameworks within which the study student interaction can be prioritized. Specifically, social constructivist theory lends credence to the notion of co-creation of meaning amongst children. In contrast to research that may prioritize one theory of learning over another, the research on classroom discussion leans on both models.

This review of literature revealed significant limitations and unfounded assumptions extant in the research on small-group literature discussions. Landmark reviews of the group discussion literature have noted that the talk of teachers and students was one of the most commonly measured aspects of discussion research (Murphy et al., 2009). It is surprising, therefore, that the majority of research stops far short of adequate analysis of these talk increases.
Although increases in talk are a fundamental aspect of a discussion approach, few scholars have analyzed variation in responses at a student level. Furthermore, few studies include analysis that controlled for the interdependency inherent in studying group dynamics (Bonito, 2002). While some qualitative investigations have noted student characteristics associated with various outcomes, no existing research has systematically investigated how changes occur in students’ talk over a significant amount of time. Even the most basic features of small-group interactions have not been studied with methodologies that account for the nested structure of group data and that track change over time (i.e., multi-level modeling). While it is true that important differences between approaches can be understood by looking at broad measures of student talk, little else can be inferred regarding change in the dynamics of the group interactions. Investigation of change in students may reveal important features of small-group discussions. Based on this review of research, a finer grained analysis of talk in small-group discussions is warranted. As a result, my study was designed to answer critical questions with regard to how student talk changes over time in small-group discussions.
CHAPTER 3: METHOD

In my study I examined how student talk changed over time during participation in Quality Talk small-group literature discussions. To this end, I conducted a secondary analysis of data collected during a large-scale investigation of the Quality Talk approach. The larger study was conducted with Institute of Educational Sciences (IES) funding over 3 years (2013-2016) across multiple school sites (Grant # IES R305A130031). The following research questions focused the methodology and goals of my study.

RQ 1: How does student talk change over time during small-group literature discussions?
RQ 2: How is text type related to student talk during small-group literature discussions?
RQ 3: What student characteristics explain change in student talk over time?
RQ 4: How does student talk change relative to other group members over time?

Participants

Participants in the larger study of Quality Talk were all fourth and fifth grade students and teachers from a mix of rural/suburban schools in the northeastern United States. The larger research study took place in three phases and participants for my study were sampled from the first phase (2013-2014). My sample consisted of two fourth grade classrooms (Class X, \( n = 17 \); Class Y, \( n = 19 \), Total \( n = 36 \)) from a semi-rural private school. State standardized test results and oral reading fluency (ORF) were approximately evenly distributed at the classroom level. Notably, the average ORF of both classes was above the 75\(^{th}\) percentile (Hasbrouck & Tindal, 2006). In terms of race, most students in these classrooms appeared white, however race data
were not collected as a part of the larger study. Similarly, economic status data were not collected from the participating students. For the whole school, approximately 30% of students were reported to receive free/reduced lunch. In terms of gender, there were slightly more female students in Class X compared to Class Y. Both participating fourth grade teachers had over 10 years of classroom experience in elementary and middle school settings when the study began, and both had at least five years of experience at the fourth grade level. Descriptive data for participants is shown in Table 1.

Table 1

<table>
<thead>
<tr>
<th>Class</th>
<th>n</th>
<th>% Female</th>
<th>Mean ORF</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>17</td>
<td>58.82</td>
<td>131.64</td>
</tr>
<tr>
<td>Y</td>
<td>18</td>
<td>50.00</td>
<td>143.22</td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td>54.28</td>
<td>137.60</td>
</tr>
</tbody>
</table>

Note. ORF = Oral Reading Fluency

Procedures

All the students in my sample participated in the first phase of a three-year investigation of the Quality Talk approach to literature discussion. The larger Quality Talk research study was approved by the IRB of both Pennsylvania State University and University of North Carolina at Chapel Hill. Implementation included ongoing professional development, coaching, and teacher reflection based on researchers’ viewing of videotaped discussions. Discourse coaches were post-doctoral and doctoral students with experience in prior trials of Quality Talk. Prior to the start of school, teachers received professional development with the principal investigator of the study and trained Quality Talk coaches. Two full-day workshops were held to introduce the
project and provide training in the Quality Talk model. The two participating fourth grade classroom teachers began implementation on the third week of the academic calendar, and held their first discussion on the fifth week of the academic calendar. Professional development was supplemented with half-day workshops monthly during the rest of the school year. Monthly workshop time was devoted to reviewing the Quality Talk model, addressing teacher concerns, and soliciting feedback for improvement. Researchers reviewed video of each professional development session to ensure fidelity with Quality Talk. Formal feedback from the participating teachers was gathered via questionnaires at the midpoint and end of the intervention. The results of these questionnaires indicated that the participating teachers strongly supported the use of Quality Talk and felt it was an appropriate intervention. Teacher feedback was instrumental in improving how the research team shared the content that was to be delivered to students as well.

In addition to didactic professional development, one of two members of the research team met with each teacher throughout the school year for discourse coaching. The coaches alternated each session, so as to ensure both fidelity with the intervention model and that there were no systematic coaching differences. These coaching sessions included personalized, face to face feedback and detailed review of video-recorded Quality Talk discussions. The teachers also assessed their own performance by reviewing discussions with a researcher developed coding tool. Both teachers indicated via their formal feedback questionnaires that their coaching sessions provided helpful feedback.

**Student experience of Quality Talk.** Students experienced Quality Talk via three elements of the intervention. First, students received researcher-prepared direct instruction about Quality Talk. These were delivered by each classroom teacher approximately every two weeks and covered an important, different discourse element to be practiced that week (e.g., uptake,
authentic questions). These lessons took the form of engaging multimedia presentations with activities and lecture. Video segments of the presentations modeled discussion features and lessons included opportunities to participate in guided practice in using discourse elements. Fidelity checks on the lessons were completed to ensure that the lessons were delivered consistently across both classrooms. Feedback from teachers on these lessons was also collected. Overall, the delivery of multimedia lessons and activities were found to be consistent and changes were noted to be integrated into other phases of the larger research study.

As another element of Quality Talk, students were also provided opportunities for written reflection before reading, after reading, before discussion, and after discussion. Students wrote in literacy journals prepared by researchers in alignment with the weekly reading assignment. The journals included space for responding to text, writing questions, making predictions, and setting goals for the upcoming discussion. In addition, one section of the journal was used to measure student perceptions of the stories in terms of both interest and emotional response.

Finally, students in both classrooms participated in teacher-facilitated, small-group discussions about an assigned story from the fourth grade basal reader. The texts used for discussion were from the fourth grade Scott Foresman Reading Street series, which was the basal text purchased and used by the school prior to the Quality Talk research study. Each week a different story was assigned to students to read at home. The following day, students took a four item multiple choice pre-test to assess if they had read the selection. In the case that a student had not read, the student was given time to read in class so as to be adequately prepared for Quality Talk discussion. Students then discussed each story in their small-groups following the Quality Talk framework for discussion. Groups were formed by the Quality Talk research team in collaboration with the teachers. In the formation of these groups, Quality Talk researchers and
the teachers endeavored to create groups that were heterogeneous in terms of ability and gender prior to any teacher recommended changes due to other factors. Students in each classroom were split into a total of six groups with five or six students each. Groups were formed with the goal that no group was all male or all female. Groups were also formed to have different ability levels in each group. The groups were formed within classrooms, thus each teacher had a total of 3 groups in their respective class.

Discussions took place in the classroom during “literacy center” time, making it possible for the teacher to give attention to the group while other students completed self-directed tasks in another part of the classroom. Each discussion began with a review of expectations which gave students a sense of the pedagogical principles of Quality Talk. These expectations included not raising hands to speak, talking one at a time, giving others time to speak, considering others ideas, giving reasons for ideas, and arguing ideas not people. The teacher then began the discussion by asking a student to share one open-ended question from their literacy journal. As a part of the Quality Talk model, teachers were asked to act as much like equal participants as possible, and thus refrain from directing the discussion too much. When discussion came to a standstill or got very off track, the teachers provided facilitation, but one goal of Quality Talk was the complete transfer of discussion to students. Figure 1 shows the frequency of teacher test questions and authentic questions. Data on teacher talk (i.e., questions, uptake, teacher moves) were collected and indicated that participating teachers implemented the Quality Talk procedures with fidelity. The trend in these data suggest that the teacher’s role became less and less prominent over time. This serves as some evidence that both the Quality Talk model was
implemented with fidelity and that students gained increasing control over the flow of discourse. The discussions were video recorded approximately every other week. This frequency differed slightly around holidays and standardized testing during the year. The recorded discussions resulted in 14 discussion occurrences across the six different discussion groups. Following approximately half of the discussions, student comprehension was assessed using a researcher-developed measure of comprehension. These assessments were administered to students as a whole group approximately every other week. The assessment consisted of two multiple-choice questions to assess basic comprehension and three open-ended short answer questions to assess high level comprehension.

**Variation in group assignment.** Group assignment was variable for the first two occasions of discussion in one of the classes. During the first two discussions in Class Y there were multiple student reassignments across all three groups. The group assignments then remained stable from the third discussion through until the end of the year. Due to the impact of
group assignment on my outcome variables, it was necessary to drop the first two weeks from analysis. Full analysis was then performed on the remaining 12 discussions that took place across the year. No significant changes in group assignment took place during the following 12 discussions, with one exception. One student switched groups with another student after the 19th week of the academic year (i.e., the seventh Quality Talk discussion) due to conflicts with another student. Data resulting from the talk of these two students was ignored. Though purposely ignoring data is not ideal, group assignment is theorized to have some influence on talk. Thus, this change in assignment would introduce non-relevant variance in the individual trajectories of these two students, making inferences from their longitudinal data less reliable. No other students dropped out of groups, nor were any new students added through the school year. One group’s transcript was lost due to technical error with the video equipment (Group 6, fifth discussion, 12th week of school) and had to be completely dropped from the analysis. The discussion groups that remained stable were created to be similar in terms of their distribution of ORF and gender following this reassignment. Two groups were majority female and one group was majority male, however, following group reassignments. One group (Group 1) had the lowest mean ORF of the groups and Group 6 had the highest mean ORF. Descriptive statistics for each group, following reassignment, are shown in Table 2.
Table 2
*Descriptive Statistics of Participants by Group*

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>% Female</th>
<th>Mean ORF</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>50.00</td>
<td>116.17</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>83.33</td>
<td>145.40</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>50.00</td>
<td>135.67</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>50.00</td>
<td>135.34</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>66.67</td>
<td>137.50</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>40.00</td>
<td>156.83</td>
</tr>
</tbody>
</table>

*Note: ORF = Oral Reading Fluency*

**Data Sources**

The data utilized for my study were gathered from various sources as a part of the larger scale Quality Talk research project. Approximately half of Quality Talk discussions were video recorded and transcribed. In addition, ORF data for each student were measured at the beginning of the year by the classroom teachers. Student gender was noted and the genre of the text was also captured. The time spent in active discussion was captured via the videos of discussions.

**Video recordings.** Video recordings were made of approximately half of the Quality Talk discussions, on 14 occasions over approximately 38 weeks of the school year. This resulted in a total of 84 videos of small-group discussions across six student groups. Each video file was professionally transcribed by a transcription service, resulting in a transcript of each occasion of discussion. Students were assigned ID numbers to identify who was speaking, and the ID number followed by a colon indicated the speaker. Transcribers noted inaudible speech by writing “(inaudible)” and also noted periods of overlapping dialog by writing “(overlapping speech).”
Each resulting document was a transcription of the entire video file, which included some talk occurring outside of the discussion (e.g., “TEACHER: Class, please be quiet, our discussion is starting”).

**Oral reading fluency.** Classroom teachers assessed each student’s ORF score at the beginning of the 2012-2013 school year. Students’ oral reading fluency in words per minute (wpm) was measured using the AIMSWEB system (www.aimsweb.com). Oral reading fluency (ORF) has long been shown to be reliably correlated to reading comprehension (Dougherty Stahl, 2009). The National Reading Panel (2000) established ORF as one of the five elements necessary for early reading proficiency. Additionally, ORF provides an easily assessable proxy measure of overall reading competence and comprehension before grade four (Dougherty Stahl, 2009). Some problems with ORF as a measure of comprehension have been noted because expectations for reading proficiency often shift towards higher-level comprehension around grade four (Dougherty Stahl, 2009). In addition, some research on ORF growth has produced evidence suggesting ORF is not as related to comprehension in intermediate grade levels as in earlier grades (Paris & Hamilton, 2009). As the students in my study were measured at the beginning of their fourth-grade year, these students were well within the age range for which ORF is a well-suited measure of reading proficiency. ORF also measures speech fluency to a certain extent, as students are reading passages aloud when ORF is measured. This speech fluency component may predict change in how students participate in discussion, with more fluent speakers potentially contributing a higher number of words due to the ease by which they talk, generally.

**Student and text characteristics.** Student gender was represented as a dichotomous variable and was analyzed as a possible explainer of change in talk production as well. Each
student’s group assignment (i.e., 1–6) was investigated as a predictor of changes in talk over time using dummy coding procedures. Likewise, text type was categorized into either informational or narrative in order to investigate and control for the variation in talk that is associated with type of story. The determination whether a text was informational or narrative was based on research by Li et al. (2014).

**Data Preparation**

In order to reduce variation that may be due to the transcribers, each video transcript was individually screened and prepared prior to quantitative analysis. Decisions regarding preparation were made on various aspects of the transcripts. First, non-word and inaudible utterances were deleted and the scope of relevant of discussion was determined following methods described here. Following this, interruptions and overlapping talk were combined to generate accurate counts of turns and words. Finally, in order to corroborate decisions made in the preparation process, 10 transcripts were randomly sampled and were prepared by an independent preparer using the procedures outlined here. The preparer was a pre-doctoral graduate student in education with experience in qualitative and quantitative research.

**Non-word and inaudible speech.** Statements marked as “inaudible” were removed. Additionally, any utterances that were either not identified or identified as spoken by a non-group participant (i.e., utterances from other students in the class) were removed. Single word utterances that were intelligible (e.g., yes, no) were counted. Single word utterances that were unintelligible, however, were removed. This included filler words such as “um” and “uh.”

**Beginning of relevant discussion.** Another type of transcript preparation took place with regard to where the analysis of relevant discussion began and ended. The beginning and ending of discussion was slightly different across transcripts due the logistics involving the video-
recording equipment. Specifically, a teacher or student initiated each video recording prior to actually beginning the discussions. No separate camera-person/observer was available to make the determination of the beginning of discussion. While this has benefits, as non-participant observers may bias the performance of the observed participants, some irrelevant talk was included in each transcript. For example, transcripts included procedural talk required to gather the children and focus their attention. The transcripts also included irrelevant side conversations between children before they began the task. As these utterances were irrelevant to my study, they were ignored in analysis. The true beginning of each discussion was determined on each transcript as the first student utterance in response to a teacher statement that discussion had begun. This usually took the form of an initial question or prompt, but was clearly identified in each transcript. The ending of discussion was determined by the last student response to a question prior to when the teacher initiated the discussion wrap up.

**Defining turns.** For the purpose of my study, the definition of turn was based on the work of Mehan (1982) and Chinn et al. (2001), both of whom followed traditions of socio-linguistic analysis. I counted a turn as the occasion of one participant speaking until they completed their utterance. This definition is well established in the literature as valid and has been used frequently because of its ease of identification and interpretation (Bonito & Hollingshead, 1997; Chinn et al., 2001; Soter et al., 2008). Edelsky (1981) noted some difficulties with this definition when attempting to capture other nuances of discussion (i.e., floor or intent). As these nuanced features of discussion were not relevant to my research questions, these difficulties posed no threat to my definition of turn. In addition, my definition of turn was technical in nature as to reduce the amount of bias introduced by researcher interpretation.
(Edelsky, 1981). To more accurately count turns, however, interruptions between students were managed in the following way.

**Interruptions.** Interruptions were handled in a manner based on work of Chinn et al. (2001). Interrupted or overlapping utterances were collapsed to give each speaker credit for its full length. In some cases, the transcriber may have indicated the start of a new turn within an initial statement, but the initial statement continues to completion. For example:

Student 1: I think the story really is about the how the lumberjack...

Student 2: I thought it was...

Student 1: wanted to get back to his dog.

Student 2: more about how tall the trees were, right?

In this example, two students spoke at the same time, but did not significantly halt each other’s talk. Utterances like these were collapsed into two separate, complete turns for analysis. This was done to more accurately reflect the length of the students’ utterances. Precedent for this method can be found in how Chinn et al. (2001) handled “simultaneous turns” (p. 392).

Interruptions also took the form of statements of agreement. In these cases, a similar procedure was used to honor the length of the primary speaker’s contribution, even if interspersed with disagreements or agreements. For example:

Student 1: The biggest issue was that the kid didn’t want to go outside because...

Student 2: No...

Student 3: Exactly!

Student 1: his dad had told him it was, like, not okay, you know?
In this example, Student 1 was counted as having one turn, containing 25 words, and Students 2 and 3 were counted as each having one turn with one word. If the initial student abandoned their thought or relinquishes the floor in response to an interruption, their utterance was counted only for the words spoken prior to the interruption. For example:

Student 1: If I had to...

Student 2: I think I would do the same thing if I was Lemmy, like, really.

Student 1: Yeah, totally, I get it.

In this example, Student 1 was counted as taking two separate turns. Though Student 1 was interrupted, it is apparent that they ceded the floor as indicated by no logical continuation of their first utterance. Unlike other examples of interruptions, Student 1’s statement does not make logical sense if combined into one utterance (i.e., “If I had to yeah, totally, I get it”).

Furthermore, although Student 1 was interrupted, their second statement is a response to the interrupter. This indicates the beginning of a new turn, and that Student 1’s first thought was abandoned prior to completion.

Following these procedures, an interrater process was undertaken to ensure reproducibility of my transcript preparation decisions. Interrater agreement on the beginning and ending of discussion was almost unanimous, with only one disagreement in the 10 randomly selected transcripts. With regard to interruptions and turns, interrater agreement was calculated by dividing the number of disagreements by the number of relevant turns in each transcript. Over all 10 randomly selected discussions, interrater agreement on interruptions and turns was found to be over 95%. Thus, it was concluded that the preparation decisions made on the remaining 74 transcripts were reliable and no further interrater procedures were conducted.

Variables
Table 3 summarizes all the relevant variables included in my analysis. Once transcripts were adequately prepared, I used computer software to efficiently count relevant variables (i.e., number of words, number of turns) in each transcript. This was facilitated by a spreadsheet program (i.e., Microsoft EXCEL) that converted text files into a row by column format. Once in this format, the program automatically separated out each student turn into a single row, with the corresponding text from that turn in an immediately adjacent column. Following this procedure, a word counting command was utilized to count the words in each cell, organized by each speaker. Manual word counts were performed on 10 randomly selected turns in each discussion in order to confirm computer analysis.

Once transcripts were prepared and the beginning of discussion was established. Then discussion duration was measured via review of each discussion video. It would be logically unsound to assume students are truly speaking more if this result is just an artifact of longer discussion. To address this, I divided the number of Words and Turns by duration in seconds and converted these measures to Words/Min and Turns/Min. This provided a mathematical control on variation due to length of discussion. The duration of discussion did not only involve student talk, however. The teachers spoke throughout the discussion, as it was a part of the Quality Talk intervention to provide teacher leadership to help stalled discussions. These teacher moves were not separated out from the total time of discussion when duration was calculated for each discussion group. Through the intervention, the participating teachers were encouraged to turn over control of discussion to students as much as possible. For this reason, it was hypothesized that teacher talk would decrease over time and not significantly impact the resulting outcome variables. Significant variation in teacher talk from group to group would introduce variance into the duration and influence the resulting ratio of Words/Min and Turns/Min. To investigate this,
teacher talk was tabulated using the same method of counting words and turns in each transcript. The only teacher talk that was included in this tabulation was teacher talk that occurred after relevant discussion had begun. Teacher talk after discussions had ended was ignored.

**Variables for research question 1.** In order to investigate changes in student talk (RQ1), talk outcome variables were calculated at all included time points. First, the number of turns each student took was counted. I also counted the number of words spoken by each student at each occasion of discussion. These two measures were used to calculate a mean length of utterance (MLU) for each student at each occasion. The MLU was calculated by dividing by the number words by number of turns, creating an average turn for each student. After MLU was calculated for each student at each time point, the words and turns totals were divided by the total duration of that discussion to create a rate of Words/Min and Turns/Min for each student at each time point. The MLU variable did not require division by duration. As MLU is the average length of utterance across all the recorded turns in a discussion, discussion duration is already accounted for and further manipulation of this variable was not warranted.

Time point was operationalized as week of the academic year. Video discussions occurred approximately every other week, however, some weeks were skipped and others were concurrent. The week indicated in Table 3 denotes the week of the school year included in this analysis as opposed to simply the occasion of discussion (e.g., “Time 1, Time 2, etc.). The time points varied based on holiday breaks, breaks for standardized testing in the school district, and other unforeseen circumstances. My coding scheme for time reflected this occasional variation in the data collection schedule. Although two Quality Talk discussions occurred prior to the week coded as “0” in table 3, the initial week of data included in analysis was coded as such in order to
aid in the interpretation of longitudinal models, but can be converted to the week of the academic
calendar by adding nine to each week.

Table 3

<table>
<thead>
<tr>
<th>Coded Week</th>
<th>Week of Academic Calendar</th>
<th>Title of Story</th>
<th>Text Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>9</td>
<td>Horse Heroes</td>
<td>Informational</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>So You Want to Be President</td>
<td>Informational</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>The Man Who Named The Clouds</td>
<td>Informational</td>
</tr>
<tr>
<td>6</td>
<td>17</td>
<td>When Night Came to The Sea</td>
<td>Narrative</td>
</tr>
<tr>
<td>10</td>
<td>19</td>
<td>Paul Bunyan</td>
<td>Narrative</td>
</tr>
<tr>
<td>12</td>
<td>21</td>
<td>Encantado: Pink Dolphins of the Amazon</td>
<td>Informational</td>
</tr>
<tr>
<td>14</td>
<td>23</td>
<td>Navajo Code Talkers</td>
<td>Informational</td>
</tr>
<tr>
<td>17</td>
<td>26</td>
<td>Smoke Jumpers</td>
<td>Informational</td>
</tr>
<tr>
<td>21</td>
<td>30</td>
<td>Cliff Hanger</td>
<td>Narrative</td>
</tr>
<tr>
<td>23</td>
<td>32</td>
<td>Moonwalk</td>
<td>Narrative</td>
</tr>
<tr>
<td>26</td>
<td>35</td>
<td>Jim Thorpe’s Bright Path</td>
<td>Informational</td>
</tr>
<tr>
<td>28</td>
<td>37</td>
<td>A Gift from the Heart</td>
<td>Narrative</td>
</tr>
</tbody>
</table>

Variables for research question 2. Other variables of interest related to variation in
student talk over time were measured. The variables relevant to RQ 1 remained unchanged and
were carried over for the purposes of addressing RQ 2 in analysis. As genre has been shown to
influence talk, the text type (i.e., narrative or informational) was coded as a dichotomous variable
at each time point to address RQ 2. These determinations were based on those used by the NAEP
framework for assessment and by Li et al. (2014) in their research on text type as it influenced
discourse during Quality Talk discussions.

Variables for research question 3. Variables for RQ 3 included student level variables
to explain variation in growth over time and in initial status. ORF and student gender were
analyzed as possible explainers of variance in initial talk and of change in talk production. Each
student’s group assignment (i.e., 1–6) was investigated as a predictor of variance in initial talk and changes in talk over time.

**Variables for research question 4.** In order to examine changes in the contribution of members, relative to the group, the distribution of words spoken by the entire group was analyzed as well (RQ 4). Specifically, a mean and standard deviation of total words spoken by each group at each time point was calculated. The total number of words spoken by across every group member was divided by the total number of turns to produce an estimate of the average length utterance of a typical group member. This average value was expected to vary for multiple reasons over time (i.e., teacher moves, duration of discussion). However, the standard deviation of the average utterance provided an estimate of the variance of number of words spoken per turn at the group level. In this way, I quantified the variance in talk at the group level, relative to itself. The standard deviations were examined to investigate any change in that variability over time, based on small-group research proposed by Weinberger and Fischer (2006).

**Data Analysis**

In order to adequately investigate student talk changes over time, a multi-level longitudinal data analysis strategy was employed. This methodology was considered appropriate to the study of small-groups due to the inherent interdependencies in small-group data (Bonito, 2002). Furthermore, as multiple waves of data were sampled across uneven time points, longitudinal multi-level modeling (LMLM) was appropriate because, unlike general linear model methods, LMLM can accommodate data structured in this way (Singer & Willett, 2003). I will briefly review the basic elements of longitudinal analysis here in order to provide background for this methodological choice.
Multi-level longitudinal analysis. Longitudinal analysis provides estimates of the shape of changes in data (e.g., linear, quadratic, etc.) within the temporal limits of the sample (Singer & Willett, 2003). At its most basic, longitudinal analysis can be conducted as long as multiple waves of data are collected using a reasonable measure of time and if the data reflect a continuous outcome that changes systematically (Singer & Willett, 2003). Multi-level modeling
(MLM) is necessary for any situation in which data may be influenced by a grouped or “nested” structure (Luke, 2004; Singer & Willett, 2003). The main advantage of MLM is that it allows researchers to understand and control for the influence of the group membership, separate from variance due to individuals (Luke, 2004). In a more typical, cross-sectional analysis, these “groups” are groups of individuals (e.g., students in classrooms, classrooms in schools). In the case of longitudinal change, multiple observations have been collected for every individual. Thus, longitudinal multi-level models treat each observation as the lowest level of data (Level 1) and “group” these observations by student (Level 2). Any further grouping of students such as classroom and district would be treated as a Level 3 or Level 4 variables, respectively. Furthermore, if there are too few classrooms or districts to maintain adequate statistical power, influence of these groups can be investigated using dummy variables at a lower level. There are various recommendations for adequate sample size to use LMLM approach (Singer & Willett, 2003). Although Singer and Willett (2003) pointed out there are no strict rules for sample size in LMLM, for my study I followed guidelines suggested by Snijders and Bosker (1999). They noted that MLM procedures can be employed as long as there are 30 or more Level 2 groups, which in my case means 30 or more students (Luke, 2004; Singer & Willett, 2003; Snijders & Bosker, 1999). Furthermore, diagnostic techniques to assess multivariate normality were utilized and departures from normality were noted prior to statistical analysis.

**Advantages of longitudinal multi-level models.** LMLM has significant advantages over other methods for observing and explaining change over time. First, in LMLM, the requirements regarding time points of data collection are highly flexible. Thus, measurements collected at unequally spaced time points do not pose a threat to the validity of the model, as they do in ordinary least squares (e.g., within-subjects ANOVA) models. Time points are treated as a
continuous predictor variable, allowing for the greater flexibility necessary to handle data from real world settings (Kwok et al., 2008; Singer & Willett, 2003). This can be contrasted with within-subjects ANOVA, for example, which requires balanced, equally spaced time points (Tabachnick & Fidell, 2007).

Furthermore, MLM flexibly handles missing data. MLM methods can accommodate inclusion of cases with only a few data points relative to other cases (Luke, 2004; Singer & Willett, 2003). Kwok et al. (2008) and Tabachnick and Fidell (2007) noted it is preferable for missing data to have a missing completely at random (MCAR) or missing at random (MAR) structure. Other methods, such as within-subjects ANOVA, require full deletion of any case that does not have complete data for each time point. This may result in significant decreases in sample size and threaten statistical power (Kwok et al., 2008; Singer & Willett, 2003; Tabachnick & Fidell, 2007).

MLM also has benefits with regard to the variance-covariance structure necessary for interpretable results. The variance-covariance structure of longitudinal data in MLM may be either restricted or free to be estimated from existing data (Kwok et al., 2008; Singer & Willett, 2003). To utilize either within-subjects ANOVA or Ordinary Least Squares regression, data must meet the assumptions of sphericity via compound symmetry (Tabachnick & Fidell, 2007). This requires not only that variances of measures at each time point to be equal, but also that the covariance between each contiguous time point be equal. In longitudinal data, however, this assumption is problematic because each individual is measured multiple times. Longitudinal data collection results in data where errors are likely to both autocorrelate and be heteroscedastic (Tabachnick & Fidell, 2007). Fortunately, MLM allows for the modeling of systematic error across time point without undue inflation of the standard errors of each parameter estimate
(Singer & Willett, 2003). As a result of these more accurate standard errors, tests of significance on parameter estimates can be made with much less concern for Type I error compared with an OLS or ANOVA approach (Tabachnick & Fidell, 2007).

Finally, another key strength of MLM is its utility with regard to covariates. MLM covariates can co-occur with the variables of interest (i.e., a time-varying covariate such as text type) (Luke, 2004; Singer & Willett, 2003). This allows researchers to observe how unstable factors influence the change within and across individuals (Singer & Willett, 2003). Other methodologies require covariates to be stable and time-invariant, such as age or gender. MLM can accommodate these as well, but allows for time-varying covariates to help explain change in the outcomes of interest.

**Analysis for research question 1.** Using measures gathered from the transcript data preparation, I investigated change in talk of each student over multiple observations. Each student was considered their own “group” (Level 2) into which the measures calculated at 12 time points were nested. With regard to RQ 1, each measure (i.e., Turns/Min, Words/Min, MLU) was regressed on time in terms of week of the school year, coded so that the first week of analyzed data was 0, to aid in interpretation. This resulted in three separate multi-level models. Each model generated intercepts and slopes for each student. The outcomes of note were the intercepts and slopes estimated from the sample data, which implies the rate of change over time and initial status of each variable.

Prior to fitting a growth model, an unconditional means model was fit and an inter-class correlation coefficient (ICC) calculated for each talk outcome. The unconditional means model was fit to determine if there is enough variation between students to warrant further
investigation. The equations used for this model in terms of MLU are shown in Equations 1 and 2.

\[
\text{Level 1: } MLU_{ij} = \pi_{0i} + \varepsilon_{ij} \tag{1}
\]
\[
\text{Level 2: } \pi_{0i} = \gamma_{00} + \zeta_{0i} \tag{2}
\]

Where:

\(MLU_{ij}\) = mean length of utterance by student \(i\) at time \(j\)

\(\pi_{0i}\) = person specific mean

\(\gamma_{00}\) = population average mean

\(\varepsilon_{ij}\) = within-person variance

\(\zeta_{0i}\) = between-person variance

Examination of variance components resulting from this model allowed me to determine whether there was statistically significant variation in each measure left unexplained. Using the variance components of the above unconditional means model, the interclass correlation coefficients (ICC) was calculated. The ICC provides a quantitative value for the magnitude of differences in student (i.e., Level 2) trajectories. The formula for the ICC is defined as:

\[
ICC = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_\varepsilon^2} \tag{3}
\]

Where:

\(\sigma_0^2\) = Between-person variance of person specific means around the grand mean

\(\sigma_\varepsilon^2\) = Within-person variance of individual around their own mean

Snijders and Bosker (1999) recommended that ICCs greater than .07 indicate enough Level 2 variance to continue MLM analysis.

In order to answer RQ 1, unconditional growth models for each outcome variable were fit. The Level 1 equations modeled the relationship of each separate outcome measure with the
predictor Week for each student (i.e., change over time). These models estimated trajectories of change over the occasions of discussion. The outcomes in this case are the intercepts and slopes, to determine if student talk did indeed grow over time. The covariance structure of these basic models was investigated to ascertain if rate of change was correlated with differences in initial status between students as well. Relationships between variables can be curvilinear, so additional terms were added to unconditional growth models where appropriate to improve model fit.

**Analysis for research question 2.** Next, I added the variable “Text Type” to control for variation due to genre. As genre was hypothesized to influence student talk, each talk outcome measure was investigated with Text Type inserted at Level 1. The variance components were investigated to ascertain if there was statistically significant amount of variance present after controlling for text type. The best fitting models were used for the next step of analysis.

**Analysis for research question 3.** My third research question concerned possible explanations of the change in talk outcomes over time. After the most parsimonious growth models were established at Level 1 and the influence of text type was controlled for, if there was still sufficient unexplained variance analysis would continue. To ensure parsimony, a model building approach was employed, building on the best growth models fit for RQs 1 and 2 (Raudenbush & Bryk, 2002). The variables of interest with regard to RQ 3 were Gender, ORF, and Group. Each variable was inserted at Level 2 to investigate relationships to initial status or rate of change of the three outcome variables observed. This procedure was repeated with the three different outcome variables (i.e., Words/Min, Turns/Min, MLU/Min).

**Analysis for research question 4.** The final research question concerned the variability of group participation. Group may have a significant influence on student talk outcomes, however, the number of groups in my sample data was insufficient to model at Level 3. To
understand group dynamics, group-level descriptive statistics were gathered and analyzed. I wanted to determine if groups became more equitable with regard to the length of utterance at each turn. If this was the case, the difference between students with the longest utterance and students with the shortest utterances should decrease. I hypothesized that the variance around the mean length of utterance calculated for the entire group would shrink if each student began to speak more similarly. If this was the case, the spread across utterance length would become smaller in each group over the weeks of intervention (Weinberger & Fischer, 2006). To analyze this phenomenon, the standard deviations of group talk at each time point were plotted and examined. This analysis was limited in that it only described trends over time. Unfortunately, due to the small sample size at this level of data (Group n = 6, Time points = 12), no inferential methods could be employed with confidence to determine if differences over time were due to chance (i.e., OLS regression). These data will be analyzed in terms of information and narrative stories, however, to investigate if any patterns exist specific to text type.

**Summary**

The analysis of student talk over the year of Quality Talk intervention required a series of preparatory steps and a clear plan for analysis. After preparing the transcripts, variables were calculated and the length of discussion was accounted for. Interrater procedures were also completed, ensuring that my decisions with regard to transcript preparation were replicable. Analysis was completed using LMLM procedures after generating necessary descriptive statistics and screening the data for necessary assumptions. The results of the analysis are reviewed in detail in Chapter 4, Results.
CHAPTER 4: RESULTS

The research questions of my study concerned hypothesized change in student talk over time during a year-long implementation of the Quality Talk small-group discussion intervention. These questions were investigated using longitudinal multilevel models (LMLM) of student talk outcomes. Prior to conducting LMLM analysis, transcripts were prepared and descriptive data were generated on the Level 1 and Level 2 variables. Descriptive analyses were conducted using R (version 3.2.1 "World-Famous Astronaut") and SPSS (version 23.0.0). Multilevel modeling was conducted using HLM 7.01 for Windows (Raudenbush, Bryk, & Congdon, 2010).

Descriptives and Missing Data

The outcome variables (MLU, Words/Min, Turns/Min) were examined for deviations from normality. Skewness and kurtosis values indicated departures from normality in MLU and in Words/Min. Analysis of histograms and density plots showed the distributions of both MLU and Words/Min to be positively skewed. The distribution of Turns/Min appeared to meet the assumption of normality based on skewness and kurtosis values. Additionally, Shapiro-Wilk tests of normality were conducted on each variable. Results of the Shapiro-Wilk tests suggested that the distributions of the three outcome variables were non-normal, though Turns/Min was found to be closer to normally distributed than the other variables. In order to continue analysis, I chose to utilize the robust standard error feature of HLM 7.01 software, which is robust with regard to violations of normality. I utilized this output for each research question in order to mitigate problems resulting from the non-normal outcome variables. Although a square root transformation was considered to compensate for the positive skewness in these outcome variables,
variables, I chose to leave them untransformed to ease interpretation (Tabachnick & Fidell, 2007). Descriptive analysis of the Level 2 predictor variables was conducted as well. Skewness and kurtosis values were calculated for the variables Gender and ORF. These results indicated that there were slight departures from normality and negative skewness in the ORF scores. Gender was found to be somewhat equally distributed. Results of these analyses are shown in Table 5. Descriptive statistics of each outcome variable separated out by Group assignment is shown in Table 6. In addition, descriptive statistics were calculated for each outcome variable at each included time point, separated by group assignment (Appendix Table A8–A10).

Table 5
Descriptives of Level 1 and Level 2 Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>M (SD)</th>
<th>Skewness (SE Skew.)</th>
<th>Kurtosis (SE Kurt.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLU</td>
<td>384</td>
<td>10.74 (4.24)</td>
<td>1.33 (.13)</td>
<td>3.87 (.25)</td>
</tr>
<tr>
<td>Words/Min</td>
<td>384</td>
<td>23.31 (13.74)</td>
<td>.84 (.13)</td>
<td>.82 (.25)</td>
</tr>
<tr>
<td>Turns/Min</td>
<td>384</td>
<td>2.21 (1.09)</td>
<td>.32 (.13)</td>
<td>-.43 (.25)</td>
</tr>
<tr>
<td>Level 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>35</td>
<td>.54 (.50)</td>
<td>-.70 (.40)</td>
<td>.06 (.78)</td>
</tr>
<tr>
<td>ORF</td>
<td>35</td>
<td>137.6 (30.45)</td>
<td>-.18 (.40)</td>
<td>-2.09 (.78)</td>
</tr>
</tbody>
</table>

Note. MLU = Mean Length of Utterance, ORF = Oral Reading Fluency

In terms of missing data, two time points were discarded due to inconsistencies in grouping. This reduced the total number of time points from 14 to 12. Though less than ideal, this reduction in time points did not significantly impact the feasibility of LMLM procedures, as the number of waves of data collection was still sufficient for LMLM (Singer & Willett, 2003). In addition, two students were assigned to different groups following the seventh discussion (i.e.,
week 19 of the academic calendar). Thus, any data associated with these two students past this point were treated as missing because of the confound due to change in group.

Table 6
Descriptives of Level 1 Variables, By Group

<table>
<thead>
<tr>
<th>Group</th>
<th>n</th>
<th>MLU M (SD)</th>
<th>Words/Min M (SD)</th>
<th>Turns/Min M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6</td>
<td>8.81 (3.78)</td>
<td>21.58 (10.37)</td>
<td>2.46 (1.06)</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>11.47 (4.59)</td>
<td>23.63 (14.90)</td>
<td>2.01 (0.97)</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>12.52 (5.35)</td>
<td>25.08 (17.98)</td>
<td>2.09 (1.23)</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>9.83 (3.11)</td>
<td>22.62 (13.38)</td>
<td>2.21 (1.05)</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>11.59 (4.43)</td>
<td>21.45 (11.54)</td>
<td>1.96 (1.09)</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>10.06 (3.83)</td>
<td>25.03 (12.91)</td>
<td>2.49 (1.02)</td>
</tr>
</tbody>
</table>

Note. MLU = Mean Length of Utterance,

Missing data that was not due to grouping reassignment was examined and did not have a clear mechanism of missingness. Finally, one transcript had to be removed from analysis due to problems with the recording equipment that could not be resolved. The transcript from the fifth discussion of Group 6 was unintelligible and no talk data could be ascertained from it. Even with this reduction in data, the percentage of missing data out of the total was very small (8.57%). 36 units were missing out of a maximum total possible 420 data units. Given the low percentage of missing data and no clear mechanism of missingness, all missing data were treated as missing at random.

Correlation and covariance matrices were generated for the Level 1 outcome variables (Table 6, Appendix Table A1). According to these analyses, there was a slightly negative relationship between Turns/Min and MLU. This was expected as it was hypothesized that the frequency of turns would go down as a student spoke for a greater length of time, per turn. Stronger positive relationships were observed between MLU and Words/Min as well as
Words/Min and Turns/Min. This was also predicted, as MLU was calculated using the total words and turns. There was no statistically significant correlation between gender and ORF.

Further exploratory analysis was necessary to investigate possible group differences between gender or group assignment in these Level 2 predictor variables. A Welch two sample $t$-test was performed. The results did not reveal any statistically significant differences on ORF scores between genders, $t(27.05) = -0.59, p > .05$, 95% CI [-28.26, 15.69]. To determine if there were statistically significant differences between groups in mean ORF, a one-way ANOVA was conducted. Prior to conducting the ANOVA, homogeneity of variance was confirmed via Levene’s Test, $F(5, 29) = .71, p > .05$. Results of the one way ANOVA indicated there were no statistically significant differences in ORF between the six groups, $F(5, 29) = 1.18, p > .05$.

Table 7

**Correlation Matrices**

<table>
<thead>
<tr>
<th>Variable</th>
<th>MLU</th>
<th>Words/Min</th>
<th>Turns/Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLU</td>
<td>1</td>
<td>.50***</td>
<td>-.09</td>
</tr>
<tr>
<td>Words/Min</td>
<td>...</td>
<td>1</td>
<td>.75***</td>
</tr>
<tr>
<td>Turns/Min</td>
<td>...</td>
<td>...</td>
<td>1</td>
</tr>
<tr>
<td>Level 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORF</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. *$p < .05$.  **$p < .01$.  ***$p < .001$.  

Teacher talk was collected from the transcripts in the same fashion as student talk. Each teacher’s total words and turns were tabulated at each time point. Total teacher words and turns at each time point are shown in Table 8.
To determine if there were statistically significant differences between groups in teacher words spoken, a one-way ANOVA was conducted. Prior to conducting the ANOVA, homogeneity of variance was confirmed via Levene’s Test, $F(5, 65) = .21, p > .96$. Results of the one way ANOVA indicated there were no statistically significant differences between the average words spoken by the teachers across six groups and over the 12 discussions, $F(5, 65) = 2.19, p > .05$. Teacher talk did appear to decline sharply following the 10th and 12th week of the academic year.

Table 8

Teacher Words/Turns By Group and Time Point

<table>
<thead>
<tr>
<th>Week</th>
<th>Class X</th>
<th></th>
<th>Class Y</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>9</td>
<td>884 / 58</td>
<td>765 / 52</td>
<td>468 / 37</td>
<td>726 / 80</td>
<td>618 / 65</td>
<td>735 / 87</td>
</tr>
<tr>
<td>10</td>
<td>1082 / 63</td>
<td>833 / 63</td>
<td>576 / 47</td>
<td>752 / 50</td>
<td>1060 / 84</td>
<td>1333 / 81</td>
</tr>
<tr>
<td>12</td>
<td>675 / 32</td>
<td>765 / 34</td>
<td>207 / 20</td>
<td>957 / 81</td>
<td>1340 / 98</td>
<td>--</td>
</tr>
<tr>
<td>15</td>
<td>745 / 52</td>
<td>932 / 81</td>
<td>456 / 33</td>
<td>296 / 29</td>
<td>369 / 37</td>
<td>316 / 27</td>
</tr>
<tr>
<td>19</td>
<td>696 / 47</td>
<td>173 / 26</td>
<td>533 / 41</td>
<td>262 / 24</td>
<td>478 / 27</td>
<td>359 / 34</td>
</tr>
<tr>
<td>21</td>
<td>481 / 45</td>
<td>511 / 45</td>
<td>596 / 51</td>
<td>486 / 53</td>
<td>657 / 66</td>
<td>280 / 29</td>
</tr>
<tr>
<td>23</td>
<td>772 / 64</td>
<td>829 / 80</td>
<td>989 / 61</td>
<td>266 / 26</td>
<td>635 / 45</td>
<td>614 / 34</td>
</tr>
<tr>
<td>26</td>
<td>528 / 36</td>
<td>729 / 43</td>
<td>634 / 30</td>
<td>520 / 48</td>
<td>288 / 33</td>
<td>587 / 74</td>
</tr>
<tr>
<td>30</td>
<td>654 / 50</td>
<td>645 / 42</td>
<td>414 / 27</td>
<td>549 / 38</td>
<td>459 / 31</td>
<td>511 / 47</td>
</tr>
<tr>
<td>32</td>
<td>1580 / 77</td>
<td>652 / 51</td>
<td>414 / 27</td>
<td>479 / 55</td>
<td>549 / 41</td>
<td>263 / 42</td>
</tr>
<tr>
<td>35</td>
<td>793 / 53</td>
<td>941 / 65</td>
<td>492 / 39</td>
<td>373 / 39</td>
<td>637 / 39</td>
<td>300 / 59</td>
</tr>
<tr>
<td>37</td>
<td>629 / 32</td>
<td>365 / 33</td>
<td>307 / 19</td>
<td>765 / 57</td>
<td>634 / 74</td>
<td>488 / 65</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean Words</th>
<th></th>
<th>Mean Words</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>793.25</td>
<td></td>
<td>(294.03)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>678.33</td>
<td></td>
<td>(229.80)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>507.17</td>
<td></td>
<td>(193.87)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>535.92</td>
<td></td>
<td>(223.76)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>643.66</td>
<td></td>
<td>(291.03)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>526</td>
<td></td>
<td>(309.96)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Mean Turns</th>
<th></th>
<th>Mean Turns</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50.75</td>
<td></td>
<td>(13.63)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>51.25</td>
<td></td>
<td>(17.84)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>36.00</td>
<td></td>
<td>(12.64)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>48.33</td>
<td></td>
<td>(18.76)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>53.33</td>
<td></td>
<td>(23.27)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>52.64</td>
<td></td>
<td>(21.66)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Week = Week of Academic Calendar
calendar. The teacher talk in Class Y, especially, was lower in week 15 than the prior weeks, across all the groups. There was variation between time points across the weeks of Quality Talk, such that no clear pattern of reduction or increase over time was observable from investigation of the teacher talk data.

**Research Question 1**

My first research question was “How does student talk change over time during small-group literature discussions?” Longitudinal multilevel modeling was conducted with 35 Level 2 units and 384 Level 1 units. All analyses were conducted with maximum likelihood estimation and robust standard errors. First, Interclass Correlation Coefficients (ICCs) were calculated for each outcome variable based on an unconditional means model (Appendix Tables A2-A4, Model0). The ICC for MLU was found to be 35.66%, for Words/Min it was 56.75%, and for Turns/Min it was 45.20%. These were well over the 7% threshold that is considered necessary to conduct LMLM analyses on each outcome variable (Snijders & Bosker, 1999). For all of the outcome variables Time was scaled such that the 9th week of the school year discussion was zero. This allowed interpretation of the intercept as the average of each outcome variable at the time point when discussion where groups were stable.

**MLU analyses.** It was predicted that over the course of the intervention students would speak longer utterances, more words, and fewer turns on average. I predicted that these relationships would be linear in nature, thus Time was entered as uncentered predictor of the outcome variable MLU. Results of a linear model of change in MLU over time did not adequately fit, however. This was determination was made by examining the variance components and the Time coefficient. It was found that the Time coefficient of a linear model of MLU was not statistically significant. Furthermore, there was no reduction in the residual
variance compared to the null model. Scatterplots of the MLU data by student were generated and examined. A possible curvilinear pattern was apparent, so a quadratic term was created by squaring the Time variable and adding the resulting Time-squared variable to the model. A quadratic model resulted in reductions in residual variance and variance components that were statistically significant. The coefficients for Time and Time-squared were still not statistically significant, however. Additional curvilinear models (i.e. cubic, exponential) were investigated. A model that included an additional time-cubed variable was determined to be the best fitting model of MLU change over time before accounting for any other factors. This was determined based on the fact that this model resulted in reductions in the amount residual variance compared to the quadratic model and also produced Time coefficients that were statistically significant (i.e., Time and Time-squared). Furthermore, the variance components of the slopes of the cubic model of MLU were statistically significant (Equations 4–8, Appendix Table A2, Model1).

\[
\text{Level 1: } MLU_{ij} = \pi_{0i} + \pi_{1i}TIME_{ij} + \pi_{2i}TIME^{2} + \pi_{3i}TIME^{3} + \varepsilon_{ij} \quad (4)
\]

\[
\text{Level 2: } \pi_{0i} = \gamma_{00} + \zeta_{0i} \quad (5)
\]

\[
\pi_{1i} = \gamma_{10} + \zeta_{1i} \quad (6)
\]

\[
\pi_{2i} = \gamma_{20} + \zeta_{2i} \quad (7)
\]

\[
\pi_{3i} = \gamma_{30} \quad (8)
\]

Although increasing the order of the model can threaten statistical power, the 12 time points present in these data were well over the minimum of five suggested for a model with this many parameters (Singer & Willett, 2003). This model included random effects for the intercept, initial slope, and the slope of the squared (i.e., quadratic) term. The cubic term was treated as a fixed effect due to non-convergence (Equation 8). The variance components of the intercept and slopes were sufficiently large to warrant further investigation of Level 2 predictors (Appendix Table
A2, Model). Based on these results, I determined that this unconditional growth model would best represent the rise and fall of the MLU over the course of the intervention. The results of this model suggest that the average slope of change in MLU before accounting for variance due to within or between student differences shifted at multiple points across the Quality Talk intervention, first decreasing, then increasing, then decreasing slightly again, on average.

**Words/Min analyses.** I predicted that Words/Min would show a linear, upward trajectory over time and that both the intercept and slope of the time coefficient would vary between students. Words/Min was initially modeled with a linear, unconditional growth model. Based on a possible curvilinear relationship suggested by my initial scan of student level scatterplots, quadratic and cubic models were also investigated. These models were compared using deviance statistics and examination of variance explained (Singer & Willett, 2003). Although the models with quadratic and cubic time variables had slight improvements in measures of model fit, the addition of these terms added to the number of total estimated parameters, resulting in reduced parsimony. Furthermore, the quadratic and cubic time variables did not result in statistically significant time coefficients (i.e., slopes). Based on these results, the linear model was determined to be the most appropriate fit to the data (Equation 9–11, Appendix Table A3, Model).

Level 1: \( WORDS/\text{MIN}_{ij} = \pi_{0i} + \pi_{1i}TIME_{ij} + \varepsilon_{ij} \) \hspace{1cm} (9)

Level 2: \( \pi_{0i} = \gamma_{00} + \zeta_{0i} \) \hspace{1cm} (10)
\( \pi_{1i} = \gamma_{10} + \zeta_{1i} \) \hspace{1cm} (11)

The slope for Words/Min, in a model without any other Level-1 variables, was not found to be statistically significant (Appendix Table A3, Model). This implied that there was no longitudinal growth in Words/Min on average over time before accounting for the variance.
introduced by other factors. The intercept and the slope had sufficiently large variance components to investigate Level 2 predictors, so these parameters were left as random effects and it was determined that analyses could continue for RQ 2 and RQ 3.

**Turns/Min analyses.** I predicted that Turns/Min would have a linear downward slope over time. Linear, quadratic, and cubic models were created and compared via deviance statistics, residual variance, and statistical significance of the time coefficients. Both the quadratic and cubic models of Turns/Min resulted in increases to the deviance statistics. In addition, the coefficients for the Time-squared and Time-cubed were not statistically significant. A linear model resulted in a statistically significant coefficient for Time and statistically significant variance at left at Level 2 (Equations 12–14, Appendix Table A4, Model1). The coefficient for Time was positive, implying an increase in Turns/Min over time. Before accounting for any other variation at Level 1 or 2, Turns/Min increased by .01 per week on average over the course of the intervention (Appendix Table A4, Model1). The variance components of both the intercept and slope were found to be statistically significantly different from zero. Based on these results, this model was determined to be the most appropriate Level-1 model to use to investigate other variables and to model Level 2 predictors.

\[ \text{Level 1: } TURNS/MIN_{ij} = \pi_{0i} + \pi_{1i}TIME_{ij} + \varepsilon_{ij} \]  
\[ \text{Level 2: } \pi_{0i} = \gamma_{00} + \zeta_{0i} \]  
\[ \pi_{1i} = \gamma_{10} + \zeta_{1i} \]  

**Research Question 2**

My second research question was “How is text type related to student talk during small-group literature discussions?” To investigate the possible effect of genre (i.e., Text Type), it was added as a time-varying Level 1 predictor variable to the model for each outcome variable. Text
Type was coded as 1 = Informational and 0 = Narrative. This variable also served to control for possible variation due to genre in the three outcome variables.

**MLU analyses.** In order to answer this question with regard to MLU, the variable Text Type was introduced to the basic unconditional growth model that was fit for RQ 1 as a fixed effect (Equation 15, 20, Appendix Table A2, Model 2). Text Type was not modeled as a random effect due to the fact that Text Type was a pre-determined function of the stories that were assigned, and I believed the relationship between Text Type and MLU would not vary across students at any given time point.

\[
\text{Level 1: } \text{MLU}_{ij} = \pi_0 + \pi_1T_{i} + \pi_2T_{i}^2 + \pi_3T_{i}^3 + \pi_4T_{i} + \varepsilon_{ij} \quad \text{(15)}
\]

\[
\text{Level 2: } \pi_0 = \gamma_0 + \zeta_0 \quad \text{(16)}
\]

\[
\pi_1 = \gamma_{10} + \zeta_{1i} \quad \text{(17)}
\]

\[
\pi_2 = \gamma_{20} + \zeta_{2i} \quad \text{(18)}
\]

\[
\pi_3 = \gamma_{30} \quad \text{(19)}
\]

\[
\pi_4 = \gamma_{40} \quad \text{(20)}
\]

As I predicted, Text Type was found to have a statistically significant relationship with MLU. On average, the MLU of students discussing informational stories was 1.06 words less than for narrative stories (Appendix Table A2, Model 2). Examination of the deviance statistic indicated that this model has a slightly better fit than the unconditional growth model of MLU. Due to the fact that the Level 1 variable Text Type was found to have statistical significance in the model of MLU over time, it was integrated into the model that was investigated to answer questions about the Level 2 predictors (RQ 3).

**Words/Min analyses.** The unconditional linear model for Words/Min was also investigated with Text Type entered as a fixed effect at Level 1. The results of this model
showed that Text Type had a statistically significant relationship with Words/Min. Analysis of this model indicated that students spoke 1.86 fewer words per minute during informational discussions. The insertion of Text Type at Level 1 controlled for any variation due to text that may have masked change in Words/Min. The slope of the Time coefficient in this model remained at zero, however, with the addition of Text Type.

A possible confound on the measurement of Words/Min was the number of Turns/Min a student took. If students took fewer turns, fewer words would be able to be measured, however, those students’ Words/Min might still change if they spoke more per turn, over time. To control for the variance in Words/Min due to variation in Turns, Turns/Min was added to the model as a grand-mean centered random effect at Level 1 (Appendix Table A3, Model2). The relationship between Words/Min and Turns/Min was found to be statistically significant, along with Text Type and Time (Equation 21). Although Turns/Min and Words/Min were strongly correlated \( (r(382) = .75, p < .001) \), the addition of Turns/Min resulted in an improvement in model fit according to deviance statistics. Although the Time coefficient was not statistically significant, but the variance components of Time and Turns/Min were statistically significant. Thus, the results indicated that both Level 1 predictors Time and Turns/Min had sufficient amounts of variance to examine Level 2 predictors (RQ3).

Level 1: \[ WORDS/MIN = \pi_{0i} + \pi_{1i}\text{TIME}_{ij} + \pi_{2i}\text{TEXT}_{ij} + \pi_{3i}\text{TURNS}/\text{MIN}_{ij} + \varepsilon_{ij} \] (21)

Level 2: \[ \pi_{0i} = \gamma_{00} + \zeta_{0i} \] (22)
\[ \pi_{1i} = \gamma_{10} + \zeta_{1i} \] (23)
\[ \pi_{2i} = \gamma_{30} \] (24)
\[ \pi_{3i} = \gamma_{40} + \zeta_{1i} \] (25)
**Turns/Min analyses.** Finally, the possible relationship of text type on Turns/Min was investigated. Text Type was included as a fixed effect at Level 1 of the unconditional growth model. Results showed that Text Type did not have a statistically significant effect. (Appendix Table A4, Model 2). This variable was dropped from further analysis and the unconditional growth model was used to investigate the unaccounted for variance still present in this outcome variable (RQ 3).

**Research Question 3**

The next research question I investigated was “What student characteristics explain change in student talk over time?” Due to the fact that there was still unaccounted for variance between students at Level 2 in each outcome variable, explanatory variables were inserted into the model. ORF, Gender, and Group were investigated for their potential relationship to the variation between students in either their initial status or their slope over time. Each of these Level 2 variables were inserted into the best fitting Level 1 models for each outcome variable.

**MLU analyses.** Level 2 predictors of variation in change in MLU were investigated using the base cubic model that controlled for within student variation due to Text Type (Equation 15). In this model, the intercept, the linear, and the quadratic term were modeled as random effects. The cubic term and Text Type were modeled as fixed effects. It was found that Gender did not have a statistically significant relationship with the variation around the intercept or the time coefficients. The variable ORF was inserted with grand-mean centering in order to better interpret the intercept of MLU when included in the model. ORF was investigated and found to have a statistically significant relationship with the initial status of MLU. For every unit increase in ORF, the MLU of students at baseline was an additional .03 above the average
ORF was not found, however, to have a statistically significant relationship with any of the random effects of interest (i.e., the time, time-squared coefficients).

Group assignment was found to have a statistically significant relationship with the variation between student trajectories, but did not have any relationship with initial status. Due to the fact that there were not enough groups to add another level of nesting to the model, Group was analyzed using dummy coding for each group. Six dummy code variables were created and used to contrast groups of students. Statistically significant differences were found between the trajectories of the groups, thus, it was necessary to contrast each group individually to ascertain the nature and structure of these differences. I found that ORF was no longer a significant predictor of variation in the intercept when the dummy coded variables were integrated into the model. Due to the complexity of the variation in curvilinear trajectories, I inspected graphs of each group contrast to aid my interpretation (Appendix Figures A1–A7). An example model of the contrast found between Group 6 and the other groups is shown in Appendix Table A5. In Appendix Table A5, Group 6 was the comparison group for all the other groups. This table shows that Group 1 had a statistically significantly different trajectory than the excluded comparison group (Group 6). This process was subsequently repeated for each group. Overall, it was found that Group 1 and 5 stood out from the other groups. Group 1 was statistically significantly different than Groups 2, 4, 5, and 6 and Group 5 was statistically significantly different from Groups 1, 2, and 3. I have summarized the results in terms of the p-values of the coefficients for each group comparison in Table 9.
Visual analysis of the graphs for the group trajectories was employed to better understand differences between groups. The MLU of Group 1 decreased until the 21st week of the academic calendar (i.e. the 8th occasion of discussion). Groups 2, 4, 5, and 6 began to increase in MLU after 12th week of the academic calendar (i.e., the 5th discussion). These groups began sloping downward around the 30th week of the academic calendar (Appendix Figures A1–A4).

Interestingly, change over time in Group 1 was not found to be statistically significantly different than Group 3 (Table 9). Analysis of the graphs of Group 5 indicated that MLU of Group 5 rose steadily over the first few included discussions, until taking a downward turn around the 26th week of the school year (Appendix Figure A5, A7). Group 5 appeared to have a parabolic trajectory, while the other groups’ trajectories had positive and negative turns over time. All of

<table>
<thead>
<tr>
<th></th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td>0.013*</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 3</td>
<td>0.09</td>
<td>0.772</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 4</td>
<td>&lt;.001***</td>
<td>0.254</td>
<td>0.2</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 5</td>
<td>&lt;.001***</td>
<td>0.039*</td>
<td>0.036*</td>
<td>0.184</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>Group 6</td>
<td>&lt;.001***</td>
<td>0.627</td>
<td>0.465</td>
<td>0.444</td>
<td>0.064</td>
<td>--</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td>0.007**</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 3</td>
<td>0.258</td>
<td>0.381</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 4</td>
<td>&lt;.001***</td>
<td>0.188</td>
<td>0.048</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 5</td>
<td>&lt;.001***</td>
<td>0.034*</td>
<td>0.01*</td>
<td>0.289</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>Group 6</td>
<td>&lt;.001***</td>
<td>0.229</td>
<td>0.058</td>
<td>0.914</td>
<td>0.253</td>
<td>--</td>
</tr>
</tbody>
</table>

Note: *p < .05. **p < .01. ***p < .001
the groups appeared to show a shortening in MLU on average during the final weeks of the intervention, however. Overall, in terms of MLU, there were statistically significant differences across the slopes and trajectories of the groups, and Group 1 and 5 stood out. With regard to what student characteristics were related to differences in change in talk over time, Gender and ORF were not statistically significant and Group assignment was.

The best fitting model for MLU over time was a cubic model that controlled for Text Type at Level 1 (Appendix Table A5). Group was found to be a statistically significant Level 2 predictor of the time and time-squared Level 1 variables (Equation 28, 29, Appendix Table A5). Following guidelines by Singer and Willet (2003), \( \text{pseudo-}R^2 \) statistics were calculated for the final model of MLU over time in order to determine how well independent variables explain the variance in MLU. The \( \text{pseudo-}R^2 \) for Level 1 residual variance was calculated by comparing the Level 1 residual variance from the null model with the Level 1 residual variance in the final model. The \( \text{pseudo-}R^2 \) at Level 1 for this model was .16 indicating that 16% of the variance within students in MLU was accounted for by the independent variables Time and Text Type. According to J. Cohen (1992) this may be considered a medium to large effect. Another \( \text{pseudo-}R^2 \) to quantify the reduction in residual variance explained by level 2 predictors was also calculated. In this case, the result of comparing the final model to the unconditional means model was negative and thus, uninterpretable. In light of this, a different \( \text{pseudo-}R^2 \) was calculated comparing the final model to the unconditional cubic growth model, in order to estimate the proportion of variance explained by the addition of Level-2 predictors to the unconditional model of change over time (Appendix Table A2, A6). The resulting \( \text{pseudo-}R^2 \) in this comparison was also negative, and thus uninterpretable. As explained by Singer and Willet (2003), in cases where the outcome variation is either exclusively between-individual or within-
individual, there may be increases in Level-2 residual variance components compared to the unconditional model. Examples of the final model equations, including dummy variables for group, are shown in Equations 26–31.

Level 1: \( MLU_{ij} = \pi_{0i} + \pi_{1i}TIME_{ij} + \pi_{2i}TIME^2 + \pi_{3i}TIME^3 + \pi_{4i}TEXT_{ij} + \varepsilon_{ij} \) \hfill (26)

Level 2: \( \pi_{0i} = \gamma_{00} + \zeta_{0i} \) \hfill (27)

\[ \pi_{1i} = \gamma_{10} + \gamma_{11}(D1) + \gamma_{12}(D2) + \gamma_{13}(D3) + \gamma_{14}(D4) + \gamma_{15}(D5) + \zeta_{1i} \] \hfill (28)

\[ \pi_{2i} = \gamma_{20} + \gamma_{21}(D1) + \gamma_{22}(D2) + \gamma_{23}(D3) + \gamma_{24}(D4) + \gamma_{25}(D5) + \zeta_{2i} \] \hfill (29)

\[ \pi_{3i} = \gamma_{30} \] \hfill (30)

\[ \pi_{4i} = \gamma_{40} \] \hfill (31)

**Words/Min analyses.** Level 2 predictors were investigated for the outcome variable Words/Min, building on the model which controlled for Turns/Min and Text Type at Level 1, with Turns/Min estimated as a random effect (Equation 21, Appendix Table A3, Model 3). Level 2 predictors were also investigated in an attempt to explain the residual variation in the relationship of Turns/Min to Words/Min. Multiple combinations of Level 2 predictor variables were attempted, including models with Level 2 predictors only on the intercept, only on Time, or only on Turns/Min. None of the Level 2 predictors modeled were found to account for the variation in how Turns/Min related to Words/Min. Gender was also not found to have any statistically significant relation with intercept or slope and thus was not included the final model. ORF was grand-mean centered and found to have a statistically significant relationship Words/Min at the intercept. These results suggested that students with higher ORF scores spoke more Words/Min. Group differences were investigated using a dummy coding procedure appropriate for analysis of categorical Level 2 predictors. This analysis revealed statistically significant differences in students’ change in Words/Min over time between groups (Table A6).
Unlike results for MLU, the associations between ORF and the intercept remained statistically significant when dummy coded variables for groups were included. Multiple contrasts were conducted by excluding each dummy coded group from the model. The results of analysis suggested that Group 6 was statistically significantly different in trajectory of Words/Min than the rest of the groups. The rest of the groups’ slopes did not statistically significantly differ in terms of slope. (Appendix Table A6, Table 10). When used as the comparison group, Group 6 appeared to have a statistically significantly more negative slope, different from the other groups (Appendix Table A6). As noted in analysis for RQ 1 and 2, the coefficient of the Time variable of Words/Min before group was added into the model was not statistically significant. This suggested that the differences in groups were masking change over time in Words/Min. As such, when separated out, the change over time in Group 6 was found to be statistically significantly different from the other groups. Using Group 6 as a comparison Group 1, 2, and 4 spoke less Words/Min over time, though with not as steep of a downward slope as Group 6. Using Group 6 as a comparison, Group 3 and 5 spoke more Words/Min over time compared to Group 6, with Group 3 having the sharpest upward slope in comparison. Groups 1–5 were not statistically significantly different from each other, unless compared with Group 6 (Table 10, Appendix Table A6). Overall, the group analysis indicated that Group 6 had a statistically significant downward slope. This appeared in spite of the lack of an increase in teacher talk (Table 8). The other group trajectories were not found to be statistically significantly different from each other and were predominantly flat.
Table 10

<table>
<thead>
<tr>
<th>Time</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td>0.608</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 3</td>
<td>0.225</td>
<td>0.426</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 4</td>
<td>0.614</td>
<td>0.984</td>
<td>0.447</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 5</td>
<td>0.259</td>
<td>0.53</td>
<td>0.782</td>
<td>0.555</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td>Group 6</td>
<td>0.024*</td>
<td>0.034*</td>
<td>0.016*</td>
<td>0.014*</td>
<td>0.005**</td>
<td>--</td>
</tr>
</tbody>
</table>

Note: *$p < .05$. **$p < .01$. ***$p < .001$.

The best fitting model of Words/Min was found to be a linear model, and included Level 1 predictors controlling for Text Type and Turns/Min. ORF was found to have a statistically significant relationship with Words/Min at the intercept. Furthermore, one group (Group 6) was found to have a statistically significantly different change in Words/Min compared with the other groups. No Level 2 predictors were found to explain the residual variation in Turns/Min at Level 1. Pseudo-$R^2$ values were calculated for Words/Min at Level 1 and Level 2. The percentage of residual variance in Words/Min that was explained by Level 1 variables (i.e., Time, Text Type, and Turns/Min) was 52.60%, which may be considered a large amount of variance according to rules of thumb outlined by J. Cohen (1992). The level 2 pseudo-$R^2$ was found to be 73.95%, which suggested a large amount of residual variance between student was explained by Level 2 predictors. With regard to the relationship between student characteristics on Words/Min, one group was found to stand out (Group 6). Statistically significant relationships were found between Words/Min, Turns/Min, and Text Type. Example final model equations for the outcome variable Words/Min are shown in Equations 32–36.

Level 1: \( WORDS/\text{MIN} = \pi_{0l} + \pi_{1l}TIME_{ij} + \pi_{2l}TEXT_{ij} + \pi_{3l}\text{TURNSTMN}_{ij} + \varepsilon_{ij} \) (32)

Level 2: \( \pi_{0l} = \gamma_{00}(ORF) + \zeta_{0l} \) (33)
\[ \pi_{1i} = \gamma_{10} + \gamma_{11}(D1) + \gamma_{12}(D2) + \gamma_{13}(D3) + \gamma_{14}(D4) + \gamma_{15}(D5) + \zeta_{1i} \]  
\[ \pi_{2i} = \gamma_{20} \]  
\[ \pi_{3i} = \gamma_{30} + \zeta_{3i} \]

**Turns/Min analyses.** Level 2 predictors were investigated in an attempt to explain variation in Turns/Min using the previously established model of change in Turns/Min over time. ORF was not found to be associated with variation in either the intercept or slope. By contrast, Gender was found to be statistically significant moderator of the intercept. Gender did not have a statistically significant relationship with the trajectory of Turns/Min (Appendix Table A7). This result indicated that at intercept, girls took .67 Turns/Min fewer than boys.

To examine the of Group, Gender was kept in the model and dummy-coded group variables were entered at Level 2. Statistically significant differences were found between some groups on both the intercept and the slope. Specifically, Group 6 was statistically significantly different from Group 2 and Group 4 at baseline. Group 2 and 4 took .88 and .49 more turns per minute than Group 6 respectively (Appendix Table A7). In addition, Group 6 was found to be statistically significantly different from Groups 1 through 5 in terms of slope. There were no statistically significant differences in the slopes between the other groups (Table 11). The trajectory of Group 6 was slightly positive in relation to the other groups. The slopes of Group 1, 2, 3, and 5 were downward trending, while Group 4 had a very slight upward slope (Appendix Table A7). Examination of the variance components of these models showed that there was no longer statistically significant variance in slope remaining after accounting for group differences. Thus, the final model for Turns/Min was a linear growth model with one predictor variable at Level 1 (Time), Level 2 predictor variables Gender and Group modeling variation on the intercept, and with Level 2 predictor variable Group modeling change over time in Turns/Min.
(Appendix Table A7). Results indicated that Group had a statistically significant relationship to intercept and slope and gender was statistically significantly associated with intercept only. A \( \text{pseudo-} R^2 \) estimating the proportion of variance within individuals explained by the independent variables in the Level 1 model was calculated to be 10%, which is considered a small to medium effect (J. Cohen, 1992).

Table 11

\textit{p-values of Group Comparisons for Turns/Min}

<table>
<thead>
<tr>
<th></th>
<th>Intercept</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td>0.673</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 3</td>
<td>0.859</td>
<td>0.491</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 4</td>
<td>0.663</td>
<td>0.208</td>
<td>0.819</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 5</td>
<td>0.843</td>
<td>0.403</td>
<td>0.998</td>
<td>0.768</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 6</td>
<td>0.109</td>
<td>0.008**</td>
<td>0.153</td>
<td>0.041*</td>
<td>0.062</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Time</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td>0.819</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 3</td>
<td>0.816</td>
<td>0.369</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 4</td>
<td>0.349</td>
<td>0.074</td>
<td>0.313</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 5</td>
<td>0.912</td>
<td>0.321</td>
<td>0.804</td>
<td>0.189</td>
<td>--</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 6</td>
<td>0.003***</td>
<td>&lt;.001***</td>
<td>&lt;.001***</td>
<td>0.01**</td>
<td>&lt;.001***</td>
<td>--</td>
<td></td>
</tr>
</tbody>
</table>

\textit{Note:} *\( p < .05 \). **\( p < .01 \). ***\( p < .001 \)

The proportion of variance between students (i.e., level 2) was found to be 27% (\textit{pseudo-} \( R^2 = .26 \)) and is considered a medium to large effect. In sum, with regard to the relationships of student characteristics on Turns/Min, results indicated that group assignment was associated with change in Turns/Min over time and gender moderated the initial number of Turns/Min taken. An example set of equations detailing the final model are shown in Equations 35–37.
Level 1: \[ \text{TURNS}_ij / \text{MIN}_{ij} = \pi_{0i} + \pi_{1i} \text{TIME}_{ij} + \epsilon_{ij} \] (35)

Level 2:
\[ \pi_{0i} = \gamma_{00}(\text{GENDER}) + \gamma_{01}(D1) + \gamma_{02}(D2) + \gamma_{03}(D3) + \gamma_{04}(D4) + \gamma_{05}(D5) + \zeta_{1i} \] (36)
\[ \pi_{1i} = \gamma_{10} + \gamma_{11}(D1) + \gamma_{12}(D2) + \gamma_{13}(D3) + \gamma_{14}(D4) + \gamma_{15}(D5) + \zeta_{1i} \] (37)

Research Question 4

Due to the fact that students in the Quality Talk study were assigned to groups that remained mostly stable over the course of the intervention, I predicted that student talk would become more similar over time in terms of the words each student spoke per turn on average. I expected to find evidence that more talkative students would talk less per turn on average and that less talkative students would talk more in a turn as the year went on. To quantify this hypothesized phenomenon, variation in talk in each group at each time point was measured and quantified by calculating the standard deviation across every utterance in each group at each time.

\[ \text{Note. Narrative texts = Week 15, 19, 30, 32, 37; Informational texts = Weeks 9, 10, 12, 21, 23, 26, 35} \]

\[ \text{Figure 2. Standard Deviation of Group Utterances} \]
point. This variation, standardized, was used as an estimate of the average spread between the students. I hypothesized that if students’ talk became more similar, there would be resulting evidence in the form of shrinking standard deviation of the average utterance of the entire group. A trend of group standard deviations becoming smaller over time would suggest that students’ talk began to match a group norm in length as the year progressed. The standard deviations of the average utterances of each group were plotted across the 12 included time points and

![Graph showing standard deviation of group utterances](image)

*Figure 3. Standard deviation of group utterances, Narrative stories*

investigated for trends. Visual analysis of the standard deviations did not reveal a positive or negative trajectory overall (Figure 2). One exception was Group 6. The standard deviations of talk for Group 6 appeared to have a downward trend, indicating less variance in the average utterance length of the group. Although the standard deviations of most groups did not become smaller as predicted, it is notable that variance appeared to increase and decrease similarly at some time points. (Figure 2). Comparing school week 10 and 12, for example, variation in every
group increased, and between week 12 and 15 all the group standard deviations decreased. These changes, however, may have been associated with differences in the type of text being discussed. In the case of the increase from week 10 to 12, both these stories were informational. Yet, in the case of the decrease from week 12 to week 15, talk about an informational story (week 12) had more variation than talk about a narrative story (week 15). To better understand how possible trends in variation were associated with text type, the data were plotted separately for narrative and informational texts. These results are shown in Figures 3 and 4. Visual analysis of the graphs suggested different trends in variation by text type. The variation in the average length of utterance of a group appeared to increase when the group was discussing narrative stories (Figure 3). Especially between week 15 and week 30, the variation between students in Groups 1 through 5 increased. The exception was Group 6. The variation between the utterances of Group 6 became smaller over time. By comparison, the variation during discussions about
informational text was relatively stable (Figure 4). A post-hoc Welch’s two sample $t$-test to determine if there were differences in the standard deviations by text type was performed. Results of this analysis did not indicate a statistically significant difference in the average standard deviations of narrative and informational text discussions, $t(63.38) = 0.81, p > .05, 95\%$ CI [-0.99, 2.36].

**Summary of Results**

Analysis of student talk during the Quality Talk intervention resulted in findings regarding the amount of talk, change in talk over time, and factors related to the change in talk. With regard to change over time (RQ 1) MLU was found to follow a cubic pattern over time, no change in trajectory was found in Words/Min, and Turns/Min rose slightly, overall. Once Text Type was controlled for (RQ 2), some differences were found between trajectories for informational vs. narrative stories. RQ 3 concerned student characteristics and group assignment. At baseline, it was found that ORF predicted differences in Words/Min and Gender predicted differences in Turns/Min. Group was found to be a significant predictor of difference in all of the outcome variables examined. The models created for these talk outcomes accounted for medium to large percentages of the overall variance in the student talk outcomes measured. Variation in each group’s average length of utterance was analyzed using descriptive statistics and trend analysis (RQ 4). Most groups’ length of turn did not appear to become more similar. Comparison of narrative and informational texts indicated turns became less similar in terms of length over time in narrative stories. Variation in group length of utterance during discussions of informational text was not markedly different over time. The standard deviation of utterances of Group 6 did appear to decline, suggesting smaller differences in length of each turn by the end of the year.
The findings here were in contrast to what I hypothesized. Briefly, it was hypothesized that MLU and Words/Min would increase over time while Turns/Min would decrease over time. I found evidence to suggest that MLU varied widely over the year of intervention and that Words/Min and Turns/Min stayed relatively stable, overall. The hypothesis that Text Type would explain differences in talk outcomes was supported by my findings. My hypothesis that gender and ORF would be associated with change in talk was only partially supported by these findings. Although gender was found to be associated with differences in Turns/Min, it was not related statistically to change in MLU or Words. With regard to ORF, these findings partially aligned with my hypotheses that ORF would be associated with more talk, as it was found to be statistically significantly related to baseline MLU and Words/Min. ORF was not, however, found to be statistically significantly associated with change in talk. In all three outcome variables studied, Group was found to have a statistically significant relationship to change over time. Finally, although I predicted that groups would speak more similarly over the course of the intervention, this prediction was not supported by these findings. My results suggest that narrative stories are associated with wider variety in length of turns over time during Quality Talk. In sum, my analysis provided a description of student talk over time that controlled for important confounds and shed light on some basic quantitative features of talk in Quality Talk small-group discussions in 4th grade. In the final chapter, I will discuss the implications of these analyses for Quality Talk, the limitations of my study, and directions for future research.
CHAPTER 5: DISCUSSION

The purpose of my study was to explore talk outcomes of fourth grade students who participated in the small-group literature discussion intervention, Quality Talk. As national educational assessment data has shown, high-level comprehension is difficult to achieve in many classrooms (National Center for Educational Statistics, 2013). Quality Talk utilizes the most effective elements of evidence-based discussion approaches to bolster high-level comprehension of text. Quality Talk integrates an ideal instructional frame for discussion, specific pedagogical principles based on research, teacher moves found to be effective, and empirically supported discourse tools to put high-level comprehension within reach for students (Wilkinson et al., 2010).

Although discussion research spans many years and studies, only a few scholars have employed quantitative methods to examine changes over time in student talk outcomes associated with small-group discussion (Li et al., 2007; Miller et al., 2014; Sun, Anderson, Lin, & Morris, 2015). My study used longitudinal multilevel modeling to generate models of how student talk changed over time in small-group literature discussions within a year of the Quality Talk intervention. Three primary outcome variables were generated and investigated. In order to organize the discussion of my findings, I will summarize the results by outcome variable, addressing my research questions using the final longitudinal models generated. Following this, I will summarize my findings regarding how groups level talk in Quality Talk. Implications for the
Quality Talk intervention and future research will be identified along with the limitations of my study and final conclusions.

**Discussion of Final LMLM Models**

**Final model of MLU.** MLU was explored using a model building LMLM approach. I first investigated change over time (RQ 1), then controlled for the Level 1 variable, Text Type (RQ 2), and next investigated Level 2 predictors of variation in MLU (RQ 3) to explore student characteristics hypothesized to explain remaining variation in MLU. The most appropriate model of MLU was found to have a curvilinear trajectory over time, first decreasing, then increasing, then decreasing again. Text Type was found to account for variation in MLU at Level 1, and group assignment was the only Level 2 predictor that was found to be highly associated with change in MLU over time.

The model generated did not align with the trajectory of MLU that I hypothesized. I hypothesized that the MLU would increase over time as students were learning to produce explanations, arguments, and other high-level critical thinking skills. Results of research from other studies of Quality Talk suggested that students participating in Quality Talk produced more frequent “elaborated explanations” over time, demonstrating evidence of higher level thinking (Li et al., 2014; Soter et al., 2008). Based on these findings, I hypothesized that students would speak more words per utterance over the course of the intervention as they utilized Quality Talk discussion skills. My findings, by contrast, showed that MLU was shorter and longer at different occasions, over time.

The differences in talk associated with text type noted by other scholars were supported by my final model of MLU (RQ 2) (Leal, 1992; Li et al., 2014). Students took longer turns, on average, when discussing narrative stories, by approximately one word per utterance. This
finding adds to the research base on how differences in genre are related to talk outcomes. Li et al. (2014) also found narrative stories elicited longer utterances and more speech in their study of Quality Talk. Their results, which were based on 10 minute samples of discussion, combined with the results of my study, provide further evidence that discussion may vary by genre. The results of the final model of MLU add to the research literature by showing that variation in talk associated with genre exists within the upwards and downwards swings in length of utterance, over time.

I found that group assignment was the only student characteristic associated with differences amongst trajectories of MLU. Trajectories of some groups were found to be statistically significantly different from each other. Group 1 and Group 5 were statistically significantly different than most other groups. Group 5 appeared to have a trajectory that was more parabolic in nature. It is unknown what characteristics of students in these groups might explain this difference. More in-depth examination of the students in each group would be necessary to answer this question. This investigation could include qualitative analysis of group dynamics, measures of engagement, and measures of interest in the story. Follow-up research could include in-depth qualitative analysis of the content of talk in Group 1 and 5. Analysis of the frequency and type of teacher moves could also shed light on these differences, following research by Lin et al. (2015). It is possible that the teacher used slightly different teacher moves in different groups, although teacher talk and turns were not found to statistically significantly differ by group (Chapter 4, Table 8). This may have an impact on the student length of utterance, as different teacher prompts affect relational thinking or evaluative thinking in different ways (Lin et al., 2015).
The results of analysis of the final model of MLU over time suggest some interesting ramifications for Quality Talk and the study of small-group discussion talk outcomes. One possible explanation for the curvilinear MLU found in these data is that length of utterance follows the implementation of the Quality Talk intervention. Average length of utterance may have shortened (i.e., Groups 1, 2, 3, 4, 6) as students were learning and practicing questioning skills in the first three occasions of discussion measured for my study. These initial reductions in MLU corresponded with mini-lessons teaching authentic, high-level thinking questions. In addition, teacher talk was found to decrease markedly following week 12 of the school year. This decrease was more dramatic in Class Y than Class X (Chapter 4, Table 8). This change in teacher talk may have been in response to discourse coaching, which encouraged teachers to allow students to self-manage discussion as much as possible. Increases in MLU that were observed around this time point may be related to the decrease in teacher talk. During this week of the school year, mini-lessons continued that covered how to answer questions using critical thinking moves such as “elaborated explanations.” The influence of the intervention may serve to explain, in part, why length of utterance began to increase, although this evidence is by and large circumstantial. The weeks that most groups were measured as having relatively long MLU were concurrent with or following the Quality Talk mini-lessons about argumentation. Students practicing these skills would be reasonably expected to speak longer per utterance as they gave reasons and evidence for their claims, as taught by the Quality Talk curriculum. As the intervention came to a close, the evidence suggested that utterances became shorter. The mini-lessons on counter-argument may explain the downturn, in part. It stands to reason that students engaged in the challenging back and forth of what Mercer (2007) called “exploratory talk” would register fewer words per turn as they argued. Utterance may have become shorter over time as
students held each other accountable to support their claims. It is also possible that the student utterances became more efficient. MLU may have decreased at the end of the intervention because students had learned how to provide reasons and evidence for a claim, without extraneous talk. Teacher talk data, though largely stable, increased during the final few weeks of the school year. These increased words spoken may have been the result of teachers utilizing more teacher moves to guide the increasingly complex arguments taking place in Quality Talk. As the teachers’ talk showed slight increases, it may be reasonable to infer that student utterances became shorter on average as they were helped in honing their counter-arguments by their teachers.

There is some precedent in the research on small-group discussions for the variability in student talk over time. In their longitudinal study of emergent leadership in small-group literature discussions, Li et al. (2007) found evidence that incidents of leadership moves developed in a curvilinear pattern over time. In another example, a study of student talk during a small-group critical thinking intervention with fourth and fifth graders revealed increases in words spoken through the sixth lesson and a downturn in the final time point (Hudgins & Edelman, 1986). Further systematic investigation of the discussion transcripts is needed to better understand the variation in MLU over time. Future researchers could utilize these quantitative results to guide a deeper examination of changes in student discussion associated with the Quality Talk intervention. Other results of research on this sample of children implementing Quality Talk has shown that Quality Talk was associated with linear increases in high-level comprehension (Murphy et al., 2016). As my study demonstrates that MLU is not linear, this suggests that MLU may be independent of comprehension performance. The model of change of length of utterance created for my study does little to indicate if utterances made were valuable to discussion,
indicated higher-level comprehension, or were off-topic. More research is needed to better understand the factors that contribute to the upward and downward swings in length of utterance, regardless, so that teachers can optimize the volume of student thinking (i.e., talk) that is on display.

**Final model of Words/Min.** The final model of Words/Min did not align with my hypothesis regarding words spoken per minute, over time. I predicted that Words/Min would have a linear upward trajectory, but the final model results suggested there to be only very slight gains in Words/Min in a minority of groups. It was hypothesized that students would speak more in the discussion as they took over more responsibility for discussion and answered each other’s questions. The results of my analysis suggested that, overall, children spoke about the same amount of words per minute throughout the year of Quality Talk intervention (RQ 1).

Similar to MLU, in the final model of Words/Min, Text Type had a statistically significant relationship with Words/Min (RQ 2). During narrative discussions, students spoke 1.86 more Words/min, even when controlling for the number of turns taken. Although this is an informative addition to the evidence about text type, the model of Words/Min still had unexplained variance between students (i.e., at Level 2). Student characteristics were investigated in an attempt to explain this variance (RQ 3). It was found that ORF was associated with how many words students spoke at the first week included in the model (i.e., baseline). As measures of ORF are at least in some part dependent on oral fluency, it stands to reason that this variable would moderate where children began in terms of their talk volume (Hasbrouck & Tindal, 2006). Broadly, it would be expected to see a correlation between words spoken per minute and the general oral fluency of a child. The reading ability of these students as measured by ORF was not found to be associated with changes in words spoken per minute, however.
The general longitudinal stability of Words/Min demonstrated in the final model suggested that the volume of student speech, when measured across all students, did not significantly change during implementation of a small-group discussion intervention designed to bolster high-level comprehension. It may be that Words/Min functions better as a simple measure of participation to determine the ratio of teacher talk to student talk. Much of the research literature reviewed for my study measured words spoken for this purpose rather than as an indicator of successful discussion (Billings & Fitzgerald, 2002; Chinn et al., 2001; Eeds & Wells, 1989; Sandora, Beck, & McKeown, 1999). When an intervention was involved, these studies were often comparing the number of words spoken in a discussion intervention to “business as usual” (Beck et al., 1996; Chinn et al., 2001; Daniels, 2002; Goatley et al., 1995; Soter et al., 2008). In prior research on discussion interventions, counting words was a way to quantify the amount of teacher control (Billings & Fitzgerald, 2002; Eeds & Wells, 1989; Hudgins & Edelman, 1986). The Words/Min variable calculated for my study was a rate of Words dependent on the duration of discussion. This was done to control for the difference in discussion length. The lack of change may be evidence of a ceiling effect for how many words a group of students can say in a certain amount of time, however (Appendix Table A10, Chapter 4, Table 8). The evidence presented here regarding the relative stability of Words/Min adds important information to the literature base on discussion interventions but also serves to highlight the limitations of simply counting words spoken during discussion.

**Final model of Turns/Min.** The analysis of Turns/Min indicated an increase in Turns/Min over time before other variables were introduced to the model. A linear upward trajectory overall suggested that students were taking more turns on average by the end of the year of Quality Talk intervention, even after controlling for the length of discussion (RQ 1). I
had hypothesized that turns would go down as student utterance became longer, however, similar to the other outcome variables, my hypothesis assumed a simplicity of discussion development that was not borne out by these data. There were a few possible explanations for the increase in Turns/Min. One possible explanation is that as the intervention went on the teachers took fewer turns, and thus students initiated speech more. This explanation is supported in part by data that indicated a reduction in teacher moves over the course of the year (Murphy et al., 2016). When tabulated, however, teacher turns did not show a systematic downward trend that might explain, to some degree, the increase in student turns (Chapter 4, Table 8). It may be that more turns were measured as an artifact of the transcript preparation process. To explain, as detailed in Chapter 3, transcripts had many incidences of overlapping speech that could not be counted towards words or turns. These overlapping speech turns were handled by removing them from analysis. As students learned better discussion behavior it is possible that more turns were intelligible (i.e., students talked over each other less), and were thus included in the collected data, creating more measured turns. Follow-up research should test this hypothesis by tabulating the number of overlapping speech incidents and analyzing these data for trends. Furthermore, there may have been a ceiling effect for Turns/Min, similar to what could be present in Words/Min. As the discussions were kept to a relatively similar length of time throughout the intervention, there may be a reasonable upper limit to how many turns are possible during the discussions.

The results of my analysis suggested that Turns/Min taken did not have a statistically significant relationship with text type, however, unlike MLU and Words/Min (RQ 2). This finding contributes to knowledge of which factors of talk are associated with differences in genre and which may be independent of genre. With regard to which student characteristics were associated with change over time (RQ 3), statistically significant relationships were found for the
variables gender and group assignment. Gender was only associated with differences in initial status, which may be expected in light of the research on gendered behaviors in small-group discussions (French & French, 1984; Howe, 1997). Group assignment, similar to the other variables analyzed, was related to differences in change in Turns/Min. Taken with the evidence regarding change over time in Words/Min, Group 6 appeared to behave differently than the other groups based on the variables measured. Although there was limited evidence from prior research regarding Turns/Min in small-group discussion interventions, this evidence adds to the findings across the other final models with regard to the importance of group assignment to talk outcomes. Further analysis is necessary to determine what characteristics of the group may explain this, as the other groups maintained general stability over time in terms of Turns/Min and more similar trajectories in Words/Min.

| Table 12 |
|------------------|------------------|------------------|
| **Summary of RQ 3 Variables** |
| MLU | Words/Min | Turns/Min |
| Baseline | Change | Baseline | Change | Baseline | Change |
| Gender | | x | | | |
| ORF | x | | x | | |
| Group | x | | x | x | x |

*Note.* “x” indicates a statistically significant relationship between the Level 2 variable and the outcome variable.

**Summary of final LMLM models.** Overall, the longitudinal multi-level models generated for my analysis revealed important facets of the talk behavior during the Quality Talk intervention. The results of these models were varied (Table 12). Although some of my hypotheses were borne out, others were not. Broadly speaking, however, longitudinal modeling revealed that text type is related to talk, and change in talk outcomes was most consistently
associated with group. The student characteristics that were investigated, although found to be associated with differences in initial status in talk outcomes, were not found to have a relationship with how student talk changed over time (Table 12). More importantly, it was found that in two of the outcomes measured, student talk did not change much over time. Variation in talk was found in how much a student spoke per utterance.

**Research Question 4**

Although LMLM models were used to model individual student change over time, I was interested in talk outcomes on the group level. Unfortunately, without a sufficient number of groups to model another level of nesting, LMLM could not be used to reliably model unexplained variance at the group level. I hypothesized that students’ talk outcomes would become more similar over time based on other research on small-group participation (Bonito & Hollingshead, 1997; Weinberger & Fischer, 2006). Only one group’s talk outcomes became more similar over time, whereas the other groups’ talk did not support my hypothesis. Most of the groups did not become more similar in terms of how much they talked in each occasion of discussion. Student words spoken per turn increased in variance when discussing narrative stories, and varied similarly over time when discussing informational stories. In light of the curvilinear pattern of MLU and the relative stability of talk outcomes over time found for RQ 1–3, this finding is not surprising. The group that had a decrease in variation, Group 6, was also the group that spoke fewer Words/Min and took more Turns/Min compared to the other groups. This data was present in spite of evidence that teacher talk in Group 6 did not appear to have a different trend from the other groups. Due to the fact that this group was talking less, the shrinking trend in standard deviations may be a mathematical artifact instead of an indication that the students began to speak more similarly.
Nonetheless, this evidence adds to the research base on student participation in small-group literature discussions. It may be the case that, even in an ideal framework for discussion, the differences in how much a child talks per utterance will generally stay stable over time. The group norms that transfer through mini-lessons may not affect the length of utterance in each discussion as much as the discussion strategies employed. Results of research into small-group literature discussions suggested that group discussions produce student leaders who direct discussions when teachers step back from their role as directors of discussion (Li et al., 2007; Sun et al., 2015). These student leaders may speak more over time as they juggle their role as both participants and leaders. This would result in less similar participation, however, the integrity of the discussion would still be present. Regardless, educators may see low participation as problematic (Carrison & Ernst-Slavit, 2005; E. Cohen & Lotan, 1995; Howe, 2013; Reznitskaya, 2012). In some experimental studies, low participators were ranked as less competent, independent of the quality of their contributions (Jaffe & Lucas, 1969; Sorrentino & Boutillier, 1975). It may be that students can gain the benefits of a small-group discussion intervention even if they contribute at highly variable rates, as other research into Quality Talk has produced promising results in terms of high-level comprehension (Li, Murphy, & Firetto, 2014; Wilkinson et al., 2010). Although, as Bonito and Hollingshead (1997) discussed, results of some research has suggested that similar participation was not always associated with effective groups. As Kuhn (2015) noted, a low quality “coalescent discussion” (p. 6) may have similar participation but limited productivity for bolstering critical thinking or problem solving. Further complicating this picture, results of other research has suggested that students rated by peers as quiet are likely to emerge as leaders (Li et al., 2007). These students may participate less frequently, but in such a way as to prompt highly productive discourse (Li et al., 2007). Future
research should continue to investigate what levels and types of participation produce high quality discussion, while taking into account variation in student characteristics.

Limitations

In my study, I attempted to address many of the problems of analysis of small-group discussion research, nonetheless, my study was limited in several ways. First, it should be noted that measuring the gross talk outcomes ignores important contextual information from discussion. Small-group literature discussions, like much in collaborative learning, are comprised of interactions between context, topic, interest, and group dynamics such as social cohesiveness (Bonito & Hollingshead, 1997; Gillies, 2014; Murphy et al., 2011; Webb & Palinscar, 1996, Slavin, 1980). These dynamics remained unmeasured in my count of words and turns.

Next, limitations were also present in the development of my outcome variables. With regard to the variable MLU, there may have been irrelevant variance present due to transcript preparation. Specifically, there were words in the transcripts that seemed to fulfill a similar function as filler words (i.e., “um” and “uh”). The transcript preparation did not remove these borderline meaningless words such as “like” and “well.” These words functioned as parts of utterances at times, and at other times appeared to be used to fill time while students composed their thoughts. These words may have added to the total utterance length of some students idiosyncratically, while not adding to other student’s utterances. As the relevance of these words is somewhat subjective, they could not be reliably removed from analysis without a more robust interrater procedure.

Inaudible statements also may have introduced irrelevant variance. Specifically, some students may have spoken more quietly than others consistently across the transcripts. This increases the likelihood of these students having their talk removed as “inaudible” and introduces
variance in MLU due to measurement error instead of true variance. In addition, data were removed in the form of each incidence of “overlapping dialog.” Each point in the transcript where there was overlapping dialog was removed because these words and turn counts would be impossible to ascertain. This is problematic, however, if certain groups were more likely to engage in overlapping dialogue than others. If some groups’ talk overlap more than others, the removal and subsequent ignoring of these data may have obscured variation that would have been informative to the MLU analysis and to group comparisons.

With regard to how the variables Words/Min and Turns/Min were calculated, there were important limitations. These variables were created as ratios to account for the difference in duration of discussion at different time points. This would make sense on its surface as a control for variance in words and turns that stemmed from discussion duration. Yet, in my preparation of these variables, the teacher talk was not removed from the total duration used as the denominator of Words/Min and Turns/Min. Teacher talk made up a significant portion of these discussions and varied from time point to time point, and across groups. It is unclear, however, just how much time was taken up by teacher talk. Without teasing apart the talk time of just the children, this method of controlling for variation in duration was highly problematic. To elaborate, the variables, as calculated for my study, did not account for variation in teacher talk. This confound calls in to question the ability for the Words/Min and Turns/Min variables to represent variance in actual student talk. Therefore, bias due to variation in time that was not due to student speech was present and not accounted for over time. Thus, the variables Words/Min and Turns/Min are suspect in terms of the quality of inference able to be made from them. There may have also been ceiling effects present due to the natural limitation on how much a child could reasonably speak per minute. In addition, on two occasions the intercom system in the classroom beeped and
discussion was temporarily interrupted while the teacher addressed the main office of the school. Although this irrelevant talk was removed, again, the time was not subtracted from calculation of total time of relevant discussion. Overall, the issues of talk included in discussion duration that were not accounted for call into question the interpretability of the findings regarding Words/Min and Turns/Min. If variables similar to these are to be accurately studied in future research, the research design should more tightly control the time of relevant discussion so there is less uncertainty as to the nature of the variance in the variables.

Limitations exist in the methods of my analysis, as well. Although the procedures I chose were thought to be reliable when employed with non-normal data, the data used in my study did exhibit some departures from normality. There were missing data due to variety of factors that could not be addressed. Some of these missing data were due to an entire transcript being uninterpretable. More important, however, are the two time points dropped from the analysis due to changes in group membership. These are important as they would have added critical information to the models of change over time. Although LMLM is robust with regard to missing data, missing the first two time points represents a significant threat to inferences possible from longitudinal data at a theoretical level. To explain, an important foundation of the inferences possible from a longitudinal study is the concept of a starting point from which change is measured. It is possible that the biggest changes occurred early in the intervention, however, those data were not available. In my study, the two-time points that were dropped from my analysis were the true baseline in my analysis, but I had to create an experimental baseline was created from the third occasion of discussion. Although done to preserve some theoretical influence of group, it sacrificed taking into account what may have been a highly influential period of discussion development. It is unknown how much overall variance over time would
differ if data from these points was included. If students talked significantly more or less during these two weeks, it is plausible that overall variance would have increased. This increase in overall variance may have led to more observable relationships between the student level characteristics and change over time. It is unknown exactly how much change was ignored and how this variation may have adjusted the results found for this study. Given that talk data does exist for these time points, follow-up analysis could include these data to explore longitudinal models of student talk, with the caveat that group assignment changed.

Other limitations were present with regard to statistical methods used. First, there were higher levels of nesting that I could not factor into these models. Group and Class could have been a Level 3 and 4, respectively. There were not enough upper level groups to achieve adequate statistical power, however, because there were only two classes and six groups. With regard to RQ 4, there was not adequate sample size to perform inferential statistical analysis on the standard deviation of the group utterances. Also, similarity of participation as measured by the standard deviation is not a reliable measure with respect to extreme values. This means that the group may appear to have less variance in the lengths of utterances simply due to reduction in participation overall (Weinberger & Fischer, 2006). Reduction in words spoken would reduce the standard deviations, but variation may still be present. This may explain how Group 6 appeared to begin to vary less, though it is unclear if this result is an artifact of fewer words and turns over time. Future research should focus on the data from Group 6, since there was evidence to suggest Group 6 differed from the other groups in a variety of ways.

Finally, the scope of my study is limited in that the outcome variables measured were not able to be statistically related to the measures of reading comprehension or markers of high-level thinking. A short statement may have more weight and insight than a longer statement, but this
was left unmeasured by my analysis. Furthermore, there is no clear evidence generated by my study that higher rates of talk are associated with the target outcomes of the Quality Talk intervention, namely high-level reading comprehension and indicators high-level thinking. Results of other research on the same sample of students is highly suggestive that increases in comprehension are associated with exposure to Quality Talk (Murphy et al., 2016). In the course of that research, evidence was gathered regarding indicators of high-level thinking via video coding the central 10 minutes of discussion. The data generated for that research indicated increases in frequency of “elaborated explanations” and “exploratory talks” that coincided with the weeks that the longitudinal model of MLU appeared to peak in this study. This coincidental evidence, while interesting, is limited in that talk outcomes were not analyzed alongside this evidence. Further research is required that methodologically links these critical factors of student talk, student comprehension of text, and student high level thinking in discussion.

**Directions for Future Research**

Future research is necessary to better understand the dynamics of small-group discussion interventions such as Quality Talk. First, similar research should be expanded to utilize control conditions. Without a control group, the strength of the evidence analyzed here regarding talk outcomes in Quality Talk is limited to sophisticated exploration and description. I am unable to make any claims about causality related to any of the underlying instructional components in Quality Talk. Future research that compares these talk variables in both a Quality Talk and a “business as usual” discussion will add to what can be inferred from longitudinal models of student talk outcomes.

With regard to measures of talk in small-groups, more variables must be investigated as possible predictors of variation between students. Interest in research studies concerning the
impact of affect in educational contexts is growing and this is a possible variable of significance with regard to talk outcomes (Linnenbrink-Garcia et al., 2011). Additional investigations of talk should also integrate variables such as interest or perceptions of status, which have been investigated in other research (Bonito & Hollingshead, 1997; Buehl, Alexander, Murphy, & Sperl, 2001; E. Cohen, 1984; Hidi, 2001; Howe, 2013). Student level characteristics like verbal comprehension as measured by brief intelligence tests could also provide interesting covariates to talk outcomes. Additionally, there are finer elements of text type that warrant inclusion in future examinations of talk outcomes, such as story structure (i.e., story, comparison, causation, etc.) and text difficulty (Li et al., 2014). Future research on small-group discussion should integrate the methods of my study with other measures (e.g., text difficulty, student interest) to control for these potential relationships with longitudinal trajectories of talk outcomes. These findings also can help to guide how discussion of text could be optimized to increase productive talk. Teachers made aware of the differences in talk about and around different text types could better attune their expectations at the student level.

**Gender and talk.** There is contrasting evidence in the research literature on gender differences in small-group literature discussions. Some researchers conducting studies in small-group literature discussions (i.e., Collaborative Reasoning) reported no evidence of gender differences in talk (Chinn et al., 2001). By contrast Evans, Alvermann, and Anders (1997) noted small-group discussions in upper elementary children reinforced sexist stereotypes, leading girls to play a “submissive role” (p. 177). More recently, scholars have suggested that small-group discussion favors some gender roles held by girls (Godinho & Shrimpton, 2003, Li et al., 2007). For example, research specific to Collaborative Reasoning found girls to be more likely to become leaders in small-group discussions (Li et al., 2007; Sun et al., 2015). The results of my
analysis suggested that in a small-group discussion framework, girls and boys spoke similarly across a variety of talk outcome variables in a peer-led dialogic discussion context (i.e., Quality Talk). This finding adds to the research on gender difference in small-group discussions (Bonito & Hollingshead, 1997; Drudy & Chathain, 2002; Evans et al., 1997; Godinho & Shrimpton, 2003; Howe, 1997; Webb & Palinscar, 1996). Based on this literature, I had hypothesized that there may have been some measurable gender differences, however, limited evidence of difference was found in my study. In the absence of a control group, it is unknown if gender differences present in a more traditional discussion situation would be present in the Quality Talk context. The gender stereotypes that affect participation may not have been fully formed in this age group (Li et al., 2007). It is also possible that by providing mini-lessons in various ideal small-group discussion elements, domination by one gender or the other was not as easily achievable as in less formal settings. Boys were found to take more turns of talk towards the beginning of the discussions included here. Thus, broadly, the results of my analysis match up with some research conducted on other small-group discussion interventions (Chinn et al., 2001; French & French, 1984; Hammersley, 1990; Howard, 1992; Webb & Kenderski, 1984). More research is needed to identify unmeasured gender differences and how gender stereotypes affect small-group discussion. There may be gender differences in the quality of contributions to discussion (i.e., leadership), as suggested by results of research on Collaborative Reasoning (Li et al., 2007; Lin et al., 2015; Sun et al., 2015). In light of the fact that gender continues to be a concern in educational interventions, future researchers must continue to work toward a better understanding of these complicated dynamics (Godinho & Shrimpton, 2003; Lloyd, 2006).

**Oral reading fluency and talk.** Results of some research has suggested that reading proficiency has a relationship to participation (i.e., talk), however, other research has suggested
that ability is not associated with quantity of student talk (Dembo & McAuliffe, 1987; Goatley et al., 1995; Webb & Kenderski, 1984). Few studies reviewed here contained analysis of longitudinal data, and few studies reported use of standardized measures such as ORF to quantify ability for analysis. The weeks of Quality Talk intervention data utilized for my study provided a unique opportunity to observe whether differences in reading ability as measured by ORF were associated with changes in talk outcomes in a small-group discussion context. This information is critical for teachers and researchers implementing this intervention.

Although my analysis identified differences between students in ORF that related to the initial status of their MLU, ORF did not relate to the trajectories of talk outcomes over time. This finding adds important information to the research literature about the relationship of ability to participation as measured by talk (Bonito & Hollingshead, 1997; Webb & Palinscar, 1996). This is important information for educators to understand as small-group literature discussion groups are formed. During a Quality Talk intervention designed for idea discussion, no evidence was found suggesting that basic reading ability of the students was related to change in talk. These results generally aligned with studies that have shown talk outcomes to be relatively independent of ability (Chinn et al., 2001; Goatley et al., 1995; Howard, 1992; Saunders & Goldenberg, 2007). It is important to note, however, that unlike some other exploratory studies with one or two groups, the evidence presented in my study controlled for the nested structure of the data and text-type, over many weeks of discussion. Methods like these are necessary to better understand grouped interventions in education (Bonito, 2002; Kuhn, 2015). Results of some research have suggested that perception of ability influences talk (Bonito & Hollingshead, 1997; E. Cohen & Lotan, 1995; Dembo & McAuliffe, 1987; Sorrentino & Boutillier, 1975). Further research is required to examine associations between perception of true ability and talk in small-groups. It is
possible that other ability expectations factor into how students talk in small-groups, or that ability expectations may have more of an impact outside the context of the intervention. Future research can potentially be improved with the increasing focus on student assessment data, nationally. When more routinely collected, comprehensive benchmark assessments will provide standardized variables to utilize in the study of how students with different skills interact with each other in small-group discussion.

**Group dynamics and talk.** The results of my analysis lend support to calls for continued research into group dynamics in classroom interventions (Bonito, 2002; E. Cohen, 1994; Howe, 2013; Kuhn, 2015). Group assignment was associated with differences in change over time and results of my analysis suggested that different groups changed in different ways. In terms of some outcomes (i.e., Words/Min and Turns/Min), a central prediction of my study was borne out by data, after pulling out the influence of group. Although mathematically obscured at first, evidence from the models accounting for group suggested that Quality Talk is associated with slight increases in Words/Min. Some groups took fewer turns and spoke more words at the end of the intervention compared to the beginning, but without comparing groups this information would not be apparent. Although only an exploratory study, the evidence across outcomes suggests that group dynamics are related to differences in student talk. Even when groups are designed to operate a similarly as possible with regard to ability levels, gender make-up, and teacher training, they may differ in trajectory from one another. Further research is required to better understand the variables that would explain these differences.

One major question centers on how well each group comprehended the story. Follow-up analysis should examine the pattern of talk in the group that achieved higher scores on comprehension compared to groups with lower comprehension scores. Scores on comprehension
assessments could also be used in follow-up research as a time-varying covariate to control for differences between students at Level-1, much like text type.

In light of this evidence, it is increasingly important to investigate group-level relationships in educational research. The evidence here shows that groups may account more for the differences in students than teachers may assume is the case when forming groups. Education researchers should further explore these dynamics by building off research on social cohesion and group function as it relates to small-group literature discussions (Slavin, 1980; Bruhn, 2014). Steps such as quantifying the group interdependence or positive social skills of group members will add critical information to how small-group discussions are implemented (Brandon & Hollingshead, 1999; Bruhn, 2014). It is possible that there are levels of group cohesion that optimize high-level critical thinking discussions. Quality Talk creates an ideal frame for discussion with a degree of social comfort from student direction, however, the argumentation skills do invite children to challenge one another to a certain extent. Future research that integrates the social cohesion and function literature will drive better strategies for implementation of Quality Talk and small-group discussion generally. In addition, more sophisticated analysis (i.e., LMLM, relevant covariates, qualitative analysis) should be applied to group research (Bonito, 2002; Singer & Willett, 2003). The differences in student talk outcomes by group can mask changes over time when the groups are averaged together. When teachers fully understand what makes an effective group, they can be intentional when strategizing how to bolster high-level thinking in small-group discussions.

Conclusions

My study of student talk outcomes over the course of a Quality Talk intervention resulted in important information regarding change in student talk and how student characteristics are
related to talk. Some evidence was found indicating significant variation over time in certain variables. Change over time occurred most dramatically in the MLU of students, which varied in a curvilinear pattern over the year. Words/Min increased and Turns/Min decreased slightly in most groups over the year. One group was measured as speaking fewer Words/Min and taking more Turns/Min over time. Text Type was controlled in these models and was found to indicate differences in volume of talk associated with genre that align with other research on this topic. Student characteristics such as gender and ORF were investigated after controlling for text type. With regard to gender, few differences were found between boys and girls. With regard to ORF, students with higher ORF spoke more words and slightly longer utterances at baseline. There were no differences in trajectories, however, associated with ORF. Finally, investigation of group assignment found that group made a significant impact on the changes in talk variables.

My study is one of a few studies of small-group talk that have analyzed a large volume of talk over time. Additionally, few have done so with longitudinal methodologies (Li et al., 2007; Lin et al., 2015). By using computer assisted transcript preparation, my study employed quantitative methods to generate nested longitudinal models of talk outcomes. Although some prior research has counted similar talk variables such as words and turns, my study was somewhat different in that I investigated how talk changed over time using methods that could account for group variation and control for confounds such as text type. These methods have been called for in the field of participation research, but are less common in the study of small-group literature discussions (Bonito, 2002). My study provided important information about how these students’ talk changed over 38 weeks; a length of time which is rare in small-group discussion research. Although my study was exploratory in nature, it represents another step towards understanding small-group talk over time.
Although the results of my analysis cast light on some features of student talk, over time, multiple questions remain. First and foremost is how the talk dynamics explored here are related to changes in comprehension. In addition, it is clear the student characteristic variables chosen for my study were not shown to be highly associated with change over time in talk. Other student-level variables should be integrated into LMLM models of talk to better explain the variation in small-groups’ talk. Furthermore, group traits show promise as a potential avenue for better understanding of important factors in small-group discussion implementation. The methods utilized for my study are promising as they can be expanded to include better level-2 predictor variables and account for group characteristics with slightly larger samples sizes. As mentioned, qualitative analysis methods should also be employed to study characteristics of groups and to better contextualize the quantitative results of LMLM models, building off the extensive group cohesiveness research literature.

Interventions such as Quality Talk have been designed to support high-level comprehension of literature, and promising evidence has been found to that end. The implication of my study, broadly, is that Quality Talk is associated with changes student talk over time as students implement new ways of talk about and around texts. Small-group research methodologies must continue to grow in sophistication, however, as intervention research continues. This research will serve to better inform the implementation of small-group discussion interventions in classrooms.
**APPENDIX: TABLES AND FIGURES**

Table A1

*Level 1 Co-variances*

<table>
<thead>
<tr>
<th>Variable</th>
<th>MLU</th>
<th>Words/Min</th>
<th>Turns/Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLU</td>
<td>17.95</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Words/Min</td>
<td>29.36</td>
<td>188.98</td>
<td>--</td>
</tr>
<tr>
<td>Turns/Min</td>
<td>-.43</td>
<td>11.23</td>
<td>1.18</td>
</tr>
<tr>
<td>Variable</td>
<td>Model₀</td>
<td></td>
<td>Model₁</td>
</tr>
<tr>
<td>------------------</td>
<td>--------</td>
<td>---------</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>Est.</td>
</tr>
<tr>
<td>Fixed Effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>10.68***</td>
<td>.46</td>
<td>10.91***</td>
</tr>
<tr>
<td>ORF</td>
<td></td>
<td></td>
<td>.03*</td>
</tr>
<tr>
<td>Time</td>
<td>-.34*</td>
<td>.15</td>
<td>-.46**</td>
</tr>
<tr>
<td>Time²</td>
<td>.04*</td>
<td>.02</td>
<td>.05**</td>
</tr>
<tr>
<td>Time³</td>
<td>&lt;.01*</td>
<td>.00</td>
<td>&lt;.01**</td>
</tr>
<tr>
<td>Text Type</td>
<td></td>
<td></td>
<td>-1.06**</td>
</tr>
<tr>
<td>Random Effects</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>2.54***</td>
<td>6.43</td>
<td>2.91***</td>
</tr>
<tr>
<td>Time</td>
<td>.34**</td>
<td>.12</td>
<td>.34**</td>
</tr>
<tr>
<td>Time²</td>
<td>&lt;.01*</td>
<td>&lt;.01</td>
<td>&lt;.01*</td>
</tr>
<tr>
<td>Within Student</td>
<td>3.41</td>
<td>3.2</td>
<td>3.17</td>
</tr>
<tr>
<td>Deviance</td>
<td>2095.73</td>
<td></td>
<td>2106.17</td>
</tr>
</tbody>
</table>

Note: Model₀ is an intercept only; Model₁ includes time, time² and time³ as a Level 1 predictors; Model₂ includes Text Type as a Level 1 predictor; Model₃ includes ORF as a Level 2 predictor, and ORF was grand-mean centered. All models estimated with 384 level-1 units and 35 level-2 units.

*p < .05. **p < .01. ***p < .001.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Model₀</th>
<th></th>
<th>Model₁</th>
<th></th>
<th>Model₂</th>
<th></th>
<th>Model₃</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>Est.</td>
<td>SE</td>
<td>Est.</td>
<td>SE</td>
<td>Est.</td>
<td>SE</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>22.88***</td>
<td>1.81</td>
<td>21.64***</td>
<td>1.79</td>
<td>24.58***</td>
<td>1.22</td>
<td>24.81***</td>
<td>1.14</td>
</tr>
<tr>
<td>ORF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.09** .03</td>
</tr>
<tr>
<td>Time</td>
<td>.09</td>
<td>.06</td>
<td>-.06</td>
<td>.04</td>
<td>-.06</td>
<td>.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text Type</td>
<td></td>
<td></td>
<td>-1.86*</td>
<td>.77</td>
<td>-1.76*</td>
<td>.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turns/Min</td>
<td></td>
<td></td>
<td>7.90***</td>
<td>.55</td>
<td>7.91***</td>
<td>.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>10.50***</td>
<td>110.27</td>
<td>9.68***</td>
<td>93.6²</td>
<td>5.99***</td>
<td>35.94</td>
<td>5.47***</td>
<td>29.93</td>
</tr>
<tr>
<td>Time</td>
<td>.237*</td>
<td>.06*</td>
<td>.17**</td>
<td>.02</td>
<td>.17**</td>
<td>.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turns/Min</td>
<td>2.00**</td>
<td>4.00</td>
<td>2.06**</td>
<td>4.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within Student</td>
<td>9.17</td>
<td>8.84</td>
<td>6.34</td>
<td>6.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deviance</td>
<td>2880.51</td>
<td></td>
<td>2875.64</td>
<td></td>
<td>2615.81</td>
<td></td>
<td>2624.11</td>
<td></td>
</tr>
</tbody>
</table>

*Note: Model₀ is intercept only. Model₁ includes time as a Level 1 predictor. Model₂ includes Text Type and Turns/Min at Level 1, Turns/Min is grand-mean centered. Model₃ includes grand mean centered ORF at Level 2. All models estimated with 384 level-1 units and 35 level-2 units.

*p < .05. **p < .01. ***p < .001.
Table A4
Multilevel Models of Turns/Min

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model₀</th>
<th>Model₁</th>
<th>Model₂</th>
<th>Model₃</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
<td>SE</td>
<td>Est.</td>
<td>SE</td>
</tr>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>2.18***</td>
<td>.13</td>
<td>1.99***</td>
<td>.13</td>
</tr>
<tr>
<td>Gender</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>.01*</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>Text Type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>.74***</td>
<td>.55</td>
<td>.66***</td>
<td>.43</td>
</tr>
<tr>
<td>Time</td>
<td>.02*</td>
<td>&lt;.01</td>
<td>.02**</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Within Student</td>
<td>.82</td>
<td>.77</td>
<td>.78</td>
<td></td>
</tr>
<tr>
<td>Deviance</td>
<td>1012.57</td>
<td>1003.37</td>
<td>1004.35</td>
<td>995.69</td>
</tr>
</tbody>
</table>

Note: Model₀ is intercept only, Model₁ includes Time as a Level 1 predictor, Model₂ includes Text Type at Level 1, Model₃ includes time at Level 1 and Gender as a Level 2 predictor. All models estimated with 384 level-1 units and 35 level-2 units, and each model had two estimated parameters. Gender was coded as 0 = Male, and 1 = Female.

*p < .05.  **p < .01.  ***p < .001.
Table A5  
*Group 6 Comparison Multilevel Model for MLU*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model\textsubscript{4}</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Est.</strong></td>
<td><strong>SE</strong></td>
</tr>
<tr>
<td><strong>Fixed Effect</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>12.06***</td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>-.44*</td>
<td>.21</td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td>-.41***</td>
<td>.10</td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td>-.08</td>
<td>.16</td>
<td></td>
</tr>
<tr>
<td>Group 3</td>
<td>-.14</td>
<td>.18</td>
<td></td>
</tr>
<tr>
<td>Group 4</td>
<td>.11</td>
<td>.14</td>
<td></td>
</tr>
<tr>
<td>Group 5</td>
<td>.37</td>
<td>.19</td>
<td></td>
</tr>
<tr>
<td>Time\textsuperscript{2}</td>
<td>.04*</td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td>.02***</td>
<td>&lt;.01</td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td>.01</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>Group 3</td>
<td>.01</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>Group 4</td>
<td>&lt;-.01</td>
<td>&lt;-.01</td>
<td></td>
</tr>
<tr>
<td>Group 5</td>
<td>-.01</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>Time\textsuperscript{3}</td>
<td>&lt;-.01**</td>
<td>&lt;.01</td>
<td></td>
</tr>
<tr>
<td>Text Type</td>
<td>-1.06**</td>
<td>.39</td>
<td></td>
</tr>
<tr>
<td><strong>Random Effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>2.93***</td>
<td>8.63</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>.25</td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>Time\textsuperscript{2}</td>
<td>&lt;.01</td>
<td>&lt;.01</td>
<td></td>
</tr>
<tr>
<td>Within Student</td>
<td>3.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Deviance</strong></td>
<td>2128.78</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note:* Model\textsubscript{4} is an example model comparing excluded, dummy-coded Group 6 compared to the other groups. This model was estimated with 384 Level-1 units, 35 Level-2 units, and seven parameters.  
*p < .05. **p < .01. ***p < .001.*
<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est.</td>
</tr>
<tr>
<td>Fixed Effect</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>24.81***</td>
</tr>
<tr>
<td>ORF</td>
<td>.08*</td>
</tr>
<tr>
<td>Time</td>
<td>-.33***</td>
</tr>
<tr>
<td>Group 1</td>
<td>.25*</td>
</tr>
<tr>
<td>Group 2</td>
<td>.27**</td>
</tr>
<tr>
<td>Group 3</td>
<td>.41**</td>
</tr>
<tr>
<td>Group 4</td>
<td>.27*</td>
</tr>
<tr>
<td>Group 5</td>
<td>.40**</td>
</tr>
<tr>
<td>Text Type</td>
<td>-1.78*</td>
</tr>
<tr>
<td>Turns/Min</td>
<td>8.17***</td>
</tr>
<tr>
<td>Random Effects</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>5.36***</td>
</tr>
<tr>
<td>Time</td>
<td>.16*</td>
</tr>
<tr>
<td>Turns/Min</td>
<td>2.06**</td>
</tr>
<tr>
<td>Within Student</td>
<td>6.31</td>
</tr>
<tr>
<td>Deviance</td>
<td>2618.91</td>
</tr>
</tbody>
</table>

*Note: Model 4 is an example model comparing excluded, dummy-coded Group 6 to the other groups. This model was estimated with 384 Level-1 units, 35 Level-2 units, and 4 parameters. Turns/Min and ORF were grand mean centered.  
*p < .05. **p < .01. ***p < .001.
### Table A7

*Group 6 Comparisons Multilevel Model for Turns/Min*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Est.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.82***</td>
<td>.16</td>
</tr>
<tr>
<td>Gender</td>
<td>-.67*</td>
<td>.25</td>
</tr>
<tr>
<td>Group 1</td>
<td>.68</td>
<td>.41</td>
</tr>
<tr>
<td>Group 2</td>
<td>.88**</td>
<td>.31</td>
</tr>
<tr>
<td>Group 3</td>
<td>.58</td>
<td>.40</td>
</tr>
<tr>
<td>Group 4</td>
<td>.49*</td>
<td>.23</td>
</tr>
<tr>
<td>Group 5</td>
<td>.59</td>
<td>.30</td>
</tr>
<tr>
<td>Time</td>
<td>.06***</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Group 1</td>
<td>-.06**</td>
<td>.02</td>
</tr>
<tr>
<td>Group 2</td>
<td>-.07***</td>
<td>.01</td>
</tr>
<tr>
<td>Group 3</td>
<td>-.06***</td>
<td>.01</td>
</tr>
<tr>
<td>Group 4</td>
<td>-.04**</td>
<td>.02</td>
</tr>
<tr>
<td>Group 5</td>
<td>-.06***</td>
<td>.01</td>
</tr>
</tbody>
</table>

**Random Effects**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Est.</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>.63***</td>
<td>.40</td>
</tr>
<tr>
<td>Time</td>
<td>.01</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Within Student</td>
<td>.77</td>
<td></td>
</tr>
</tbody>
</table>

Deviance: 1009.88

*Note: Model 4 is an example model comparing dummy coded Group 6 compared to the other groups. This model estimated with 384 Level-1 units, 35 Level-2 units, and 4 parameters.*

*p < .05.  **p < .01.  ***p < .001."
Table A8

*MLU By Group and Time Point*

<table>
<thead>
<tr>
<th>Week</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Class X, M (SD)</td>
<td>Class Y, M (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>10.02 (2.58)</td>
<td>8.90 (2.68)</td>
<td>10.75 (5.98)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.17 (4.93)</td>
<td>10.17 (1.88)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.91 (3.78)</td>
<td>11.43 (5.00)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>8.49 (1.63)</td>
<td>10.32 (2.37)</td>
<td>11.84 (5.80)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.13 (5.00)</td>
<td>8.02 (1.67)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.53 (3.94)</td>
<td>11.43 (1.67)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>9.72 (3.53)</td>
<td>7.70 (3.08)</td>
<td>10.34 (5.80)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.47 (5.58)</td>
<td>10.34 (3.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.97 (5.88)</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>8.61 (1.64)</td>
<td>10.36 (3.33)</td>
<td>13.10 (2.32)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15.69 (1.29)</td>
<td>11.30 (1.21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.33 (3.98)</td>
<td>10.63 (4.29)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>2.45 (2.45)</td>
<td>7.40 (3.12)</td>
<td>8.93 (2.61)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.82 (1.72)</td>
<td>10.37 (2.59)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.32 (1.72)</td>
<td>9.17 (2.17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>7.40 (3.53)</td>
<td>2.40 (5.69)</td>
<td>9.51 (2.87)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.37 (5.14)</td>
<td>12.10 (3.33)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.17 (1.72)</td>
<td>13.10 (2.32)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>7.88 (1.01)</td>
<td>11.90 (3.33)</td>
<td>11.52 (1.21)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.71 (3.36)</td>
<td>11.30 (2.81)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8.61 (2.46)</td>
<td>10.63 (3.62)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>7.77 (7.77)</td>
<td>8.93 (7.77)</td>
<td>11.52 (9.58)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>9.15 (3.97)</td>
<td>8.20 (2.61)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13.95 (5.18)</td>
<td>9.58 (3.62)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>7.67 (0.92)</td>
<td>7.40 (7.77)</td>
<td>9.98 (3.70)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.32 (3.97)</td>
<td>13.13 (4.87)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19.91 (11.18)</td>
<td>14.49 (2.98)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>8.24 (1.38)</td>
<td>11.32 (1.78)</td>
<td>8.46 (2.99)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.44 (1.78)</td>
<td>11.88 (4.91)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.59 (11.88)</td>
<td>14.49 (4.62)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>1.92 (1.92)</td>
<td>8.24 (1.92)</td>
<td>7.98 (3.08)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.17 (5.17)</td>
<td>11.19 (4.07)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.75 (4.75)</td>
<td>20.63 (2.02)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>7.67 (8.33)</td>
<td>8.32 (8.33)</td>
<td>7.90 (7.29)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10.77 (3.88)</td>
<td>12.25 (4.28)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>12.35 (4.49)</td>
<td>14.29 (2.59)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.00 (4.49)</td>
<td>7.63 (7.29)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Week = Week of Academic Calendar
<table>
<thead>
<tr>
<th>Week</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>19.54</td>
<td>22.39</td>
<td>21.72</td>
<td>20.03</td>
<td>24.06</td>
<td>25.15</td>
</tr>
<tr>
<td></td>
<td>(11.96)</td>
<td>(14.22)</td>
<td>(16.96)</td>
<td>(12.14)</td>
<td>(15.49)</td>
<td>(22.68)</td>
</tr>
<tr>
<td>10</td>
<td>18.21</td>
<td>29.89</td>
<td>23.19</td>
<td>17.75</td>
<td>17.37</td>
<td>21.04</td>
</tr>
<tr>
<td>12</td>
<td>21.74</td>
<td>21.25</td>
<td>26.01</td>
<td>17.24</td>
<td>16.10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(18.00)</td>
<td>(11.47)</td>
<td>(23.49)</td>
<td>(13.95)</td>
<td>(9.49)</td>
<td>--</td>
</tr>
<tr>
<td>15</td>
<td>17.96</td>
<td>22.91</td>
<td>29.23</td>
<td>24.91</td>
<td>23.25</td>
<td>24.16</td>
</tr>
<tr>
<td></td>
<td>(11.81)</td>
<td>(11.93)</td>
<td>(12.02)</td>
<td>(12.08)</td>
<td>(9.03)</td>
<td>(10.21)</td>
</tr>
<tr>
<td>19</td>
<td>20.01</td>
<td>24.73</td>
<td>26.15</td>
<td>20.07</td>
<td>20.26</td>
<td>23.81</td>
</tr>
<tr>
<td></td>
<td>(11.87)</td>
<td>(2.25)</td>
<td>(16.77)</td>
<td>(11.20)</td>
<td>(12.17)</td>
<td>(10.77)</td>
</tr>
<tr>
<td>21</td>
<td>24.00</td>
<td>22.60</td>
<td>26.97</td>
<td>22.24</td>
<td>23.88</td>
<td>23.51</td>
</tr>
<tr>
<td></td>
<td>(5.90)</td>
<td>(13.86)</td>
<td>(16.50)</td>
<td>(14.44)</td>
<td>(18.65)</td>
<td>(10.15)</td>
</tr>
<tr>
<td></td>
<td>(8.45)</td>
<td>(10.02)</td>
<td>(16.69)</td>
<td>(17.36)</td>
<td>(9.55)</td>
<td>(12.23)</td>
</tr>
<tr>
<td>26</td>
<td>25.21</td>
<td>23.57</td>
<td>25.47</td>
<td>28.99</td>
<td>23.86</td>
<td>26.31</td>
</tr>
<tr>
<td></td>
<td>(5.37)</td>
<td>(12.31)</td>
<td>(18.48)</td>
<td>(22.50)</td>
<td>(12.57)</td>
<td>(17.30)</td>
</tr>
<tr>
<td>30</td>
<td>26.98</td>
<td>19.59</td>
<td>21.03</td>
<td>23.88</td>
<td>22.25</td>
<td>24.83</td>
</tr>
<tr>
<td></td>
<td>(8.94)</td>
<td>(3.61)</td>
<td>(19.29)</td>
<td>(9.68)</td>
<td>(10.01)</td>
<td>(10.93)</td>
</tr>
<tr>
<td>32</td>
<td>19.57</td>
<td>32.90</td>
<td>25.55</td>
<td>32.49</td>
<td>29.13</td>
<td>30.09</td>
</tr>
<tr>
<td></td>
<td>(6.92)</td>
<td>(28.70)</td>
<td>(23.87)</td>
<td>(12.99)</td>
<td>(12.18)</td>
<td>(9.29)</td>
</tr>
<tr>
<td></td>
<td>(3.56)</td>
<td>(12.29)</td>
<td>(27.84)</td>
<td>(18.09)</td>
<td>(12.99)</td>
<td>(9.87)</td>
</tr>
<tr>
<td>37</td>
<td>26.88</td>
<td>25.06</td>
<td>26.81</td>
<td>19.48</td>
<td>23.68</td>
<td>26.33</td>
</tr>
<tr>
<td></td>
<td>(12.88)</td>
<td>(21.40)</td>
<td>(22.06)</td>
<td>(13.62)</td>
<td>(10.68)</td>
<td>(14.20)</td>
</tr>
</tbody>
</table>

*Note.* Week = Week of Academic Calendar
Table A10

**Turns/Min By Group and Time Point**

<table>
<thead>
<tr>
<th>Week</th>
<th>Group</th>
<th>Class X, M (SD)</th>
<th>Class Y, M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>2.15 (1.57)</td>
<td>2.14 (1.25)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.46 (0.59)</td>
<td>2.88 (1.49)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.41 (0.88)</td>
<td>1.94 (0.86)</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>2.01 (1.18)</td>
<td>1.63 (0.24)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.91 (1.15)</td>
<td>2.48 (0.88)</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>1.85 (1.03)</td>
<td>3.08 (1.06)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.48 (0.88)</td>
<td>2.46 (1.40)</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>2.25 (1.19)</td>
<td>1.62 (0.25)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.46 (1.40)</td>
<td>2.69 (1.55)</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>3.23 (0.71)</td>
<td>2.13 (0.50)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.94 (1.55)</td>
<td>2.69 (1.55)</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>2.64 (1.06)</td>
<td>2.01 (0.55)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.69 (1.59)</td>
<td>2.69 (1.56)</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>3.23 (0.45)</td>
<td>2.52 (0.64)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.99 (1.48)</td>
<td>2.94 (1.92)</td>
</tr>
<tr>
<td>26</td>
<td></td>
<td>3.40 (0.71)</td>
<td>1.76 (0.43)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.52 (0.82)</td>
<td>1.92 (1.36)</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>2.39 (0.83)</td>
<td>2.78 (1.04)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.92 (1.36)</td>
<td>1.94 (1.36)</td>
</tr>
<tr>
<td>32</td>
<td></td>
<td>3.03 (0.40)</td>
<td>1.03 (0.74)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.94 (1.40)</td>
<td>1.94 (1.40)</td>
</tr>
<tr>
<td>35</td>
<td></td>
<td>2.22 (0.55)</td>
<td>1.76 (1.18)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.85 (1.16)</td>
<td>1.85 (1.16)</td>
</tr>
<tr>
<td>37</td>
<td></td>
<td>2.22 (0.55)</td>
<td>1.76 (1.18)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.85 (1.16)</td>
<td>1.85 (1.16)</td>
</tr>
</tbody>
</table>

**Note.** Week = Week of Academic Calendar
Figure A1. Comparison of MLU, Group 2 (Red) to Group 1 (Blue)
Note. X Axis value is converted to the week of the academic calendar by adding 9.

Figure A2. Comparison of MLU, Group 4 (Red) to Group 1 (Blue)
Note. X Axis value is converted to the week of the academic calendar by adding 9.
Figure A3. Comparison of MLU, Group 5 (Red) to Group 1 (Blue)
Note. X Axis value is converted to the week of the academic calendar by adding 9

Figure A4. Comparison of MLU, Group 6 (Red) to Group 1 (Blue)
Note. X Axis value is converted to the week of the academic calendar by adding 9
Figure A5. Comparison of MLU, Group 5 (Red) to Group 2 (Blue)
Note. X Axis value is converted to the week of the academic calendar by adding 9

Figure A6. Comparison of MLU, Group 4 (Red) to Group 3 (Blue)
Note. X Axis value is converted to the week of the academic calendar by adding 9
Figure A7. Comparison of MLU, Group 5 (Red) to Group 3 (Blue)
Note. X Axis value is converted to the week of the academic calendar by adding 9

Figure A8. Individual Student Trajectories of Final MLU Model.
Note. X Axis value is converted to the week of the academic calendar by adding 9
REFERENCES


Cohen, E. (1994). Restructuring the classroom: conditions for productive small groups. Review


Nyikos, M., & Hashimoto, R. (1997). Constructivism theory applied to collaboration learning in


